(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/10658

(22) International Filing Date: 2 April 2001 (02.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/193,839   31 March 2000 (31.03.2000)   US

(71) Applicant: UNIVERSITY OF SOUTHERN CALIFORNIA [US/US]; 3716 Hope Street, Los Angeles, CA 98007-4344 (US).

(72) Inventors: LAIRD, Peter; 649 Forrest Avenue, South Pasadena, CA 91010 (US). EADS, Cindy; 635 Prospect Street #B15, South Pasadena, CA 91010 (US).

(74) Agent: DAVISON, Barry, L.; Davis Wright Tremaine, LLP, 2600 Century Square, 1501 Fourth Avenue, Seattle, WA 98101-1688 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
—   with international search report
—   before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: EPIGENETIC SEQUENCES FOR ESOPHAGEAL ADENOCARCINOMA

(57) Abstract: There is disclosed a diagnostic or prognostic assay for cancer, particularly gastrointestinal and esophageal adenocarcinoma. Specifically, the present invention provides a methylation pattern that can be assayed by standard methylation assays of CpG islands, including which genes are hypermethylated and which genes are unmethylated in gastrointestinal and esophageal adenocarcinomas, Barrett's esophagous, and normal squamous mucosa.

EPIGENETIC SEQUENCES FOR ESOPHAGEAL ADENOCARCINOMA

CROSS-REFERENCE TO RELATED APPLICATIONS

5   This application claims priority to United States *Provisional* Patent Application
Serial No. 06/193,839, entitled EPIGENETIC SEQUENCES FOR ESOPHAGEAL
ADENOCARCINOMA, filed 31 March 2000.

STATEMENT REGARDING FEDERALLY-SPONSORED RESEARCH

10   This work was supported by NIH/NCI grant R01 CA 75090 to P.W.L. The United
States has certain rights in this invention, pursuant to 35 U.S.C. § 202(c)(6).

TECHNICAL FIELD OF THE INVENTION

           The present invention provides a diagnostic or prognostic assay for gastrointestinal
15   adenocarcinoma, and particularly esophageal adenocarcinoma ("EAC"). Specifically, the
present invention provides a multi-geneic *epigenetic fingerprint* or methylation pattern,
that can be assayed by standard methylation assays of CpG island methylation status, and
that comprises the relative methylation status of two or more genes in gastrointestinal
carcinomas, normal squamous cells, and EAC.

20

BACKGROUND OF THE INVENTION

           DNA methylation and cancer. DNA methylation patterns are frequently altered in
human cancers. These methylation changes include genome-wide hypomethylation as
well as regional hypermethylation (Jones & Laird, *Nat Genet.* 21:163-167, 1999).
25   Aberrant hypermethylation in cancer cells often occurs at CpG islands, which are
generally protected from methylation in normal tissues. Hypermethylation of promoter
CpG islands (that is, CpG islands located in promoter regions of genes) has been
associated with transcriptional silencing in many types of human cancers.

           Methylation patterns of genes can provide different types of useful information
30   about a cancer cell. First, each tumor type (*i.e.*, breast, colon, esophagus, etc.) has a
characteristic set of genes with an increased propensity to become methylated (Costello et
al., *Nat. Genet.* 24:132-138, 2000). For example, *RB1* is known to be hypermethylated in
retinoblastoma (Stirzaker et al., *Cancer Res.* 57:2229-2237, 1997; Sakai et al., *Am. J.*
Hum. Genet. 48:880-888, 1991), but not in acute myelogenous leukemia (Kornblau & Qiu,
35   *Leuk. Lymphoma.* 35:283-288, 1999; Melki et al., *Cancer Res.* 59:3730-3740, 1999).

           Second, an individual tumor within a single patient has a unique epigenetic
fingerprint reflective of the evolution of that tumor as compared to a tumor of the same
type in a different patient (Costello et al., *Nat. Genet.* 24:132-138, 2000).

           Generally, however, most studies of epigenetic alterations in cancer have focused
40   primarily on either a very small set of known genes (Jones & Laird, *Nat Genet.* 21:163-

167, 1999; Baylin & Herman, *Trends Genet.* 16:168-174, 2000) or on the global analysis of unknown CpG islands (Costello et al., *Nat. Genet.* 24:132-138, 2000), and thus do not provide a suitable diagnostic and/or prognostic framework.

5   *Esophageal adenocarcinoma ("EAC").* Esophageal adenocarcinoma ("EAC") arises from a multistep process whereby normal squamous mucosa undergoes metaplasia to specialized columnar epithelium (Intestinal Metaplasia (IM) or Barrett's esophagus), which then ultimately progresses to dysplasia and subsequent malignancy (Barrett et al., *Nat. Genet.* 22:106-109, 1999; Zhuang et al., *Cancer Res.* 56:1961-4, 1996). The incidence of EAC has increased rapidly in the Western World over the past three decades
10   (Devesa et al., *Cancer.* 83:2049-2053, 1998; Jankowski et al., *Am. J. Pathol.* 154:965-973, 1999).

Unfortunately, epigenetic studies of this model have so far been limited to the DNA methylation analysis of a few genes (Wong et al., *Cancer Res.* 57:2619-2622, 1997; Klump et al., *Gastroenterology.* 115:1381-1386, 1998; Eads et al., *Cancer Res.* 60:5021-
15   5026, 2000).

CpG island methylator phenotype ("CIMP"). It has previously been reported that a subset of colorectal and gastric tumors display a CpG island methylator phenotype ("CIMP"), characterized by widespread, aberrant hypermethylation changes affecting multiple loci in a single tumor (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686,
20   1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999). This is reflected in a bimodal distribution of the frequency of the number of genes methylated in a group of tumors (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686, 1999). CIMP tumors are a distinct group of tumors that are defined by a high degree of concordant CpG island hypermethylation of genes exclusively methylated in cancer, or type C genes. CIMP is
25   now thought to be a new, distinct, yet major pathway of tumorigenesis (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686, 1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999).

However, the role, if any, of the CIMP pathway in the tumor evolution of EAC is still uncharacterized, because the previous epigenetic studies only analyzed one (Wong et
30   al., *Cancer Res.* 57:2619-2622, 1997; Klump et al., *Gastroenterology.* 115:1381-1386, 1998) or a few genes (Eads et al., *Cancer Res.* 60:5021-5026, 2000).

Therefore, there is a need in the art for novel methods of cancer detection, chemoprediction and prognostics. There is a need in the art to define novel coordinate patterns of CpG island methylation changes at multiple loci during different steps of a
35   disease, such as cancer. There is a need in the art to determine tumor-type-specific, and patient-specific epigenetic patterns or fingerprints. There is a need in the art to provide biomarkers or probes, such as EAC-specific biomarkers or probes, that can be used in diagnostic and/or prognostic methods for the treatment of cancer. There is a need in the art to determine whether esophageal adenocarcinoma displays a CIMP. There is a need in

2

the art for novel methods for determining the stage of a tumor. The present invention addresses these needs.

SUMMARY OF THE INVENTION

5   The present invention provides a method for diagnosing cancer or cancer-related conditions from tissue samples, comprising: (a) obtaining a tissue sample from a test tissue or region to be diagnosed; (b) performing a methylation assay of the tissue sample, wherein the methylation assay determines the methylation state of genomic CpG sequences, wherein the genomic CpG sequences are located within at least one gene

10   sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR*, and combinations thereof; and (c) making a diagnostic or prognostic prediction of the cancer based, at least in part, upon the methylation state of the genomic CpG sequences. Preferably, the genomic CpG sequences

15   located within at least one gene sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2* and *TYMS*, correspond to genomic CpG sequences of CpG islands. Preferably, the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3,*

20   *CTNNB1, PTGS2, TYMS* and *MTHFR* gene sequences are those defined by the specific oligonucleotide primers and probes corresponding to SEQ ID Nos:1-60, 64 and 65, as listed in TABLE II, or portions thereof. Preferably, the CpG islands are located within the promoter regions of the genes. Preferably, the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3,*

25   *CTNNB1, PTGS2,* and *TYMS* gene sequences correspond to any CpG island sequences associated with the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID Nos:1-54, 58-60, 64 and 65, as listed in TABLE II, or portions thereof, wherein the associated CpG island sequences are those contiguous sequences of genomic DNA that encompass at least one nucleotide of the sequences defined by the

30   specific oligonucleotide primers and probes corresponding to SEQ ID Nos:1-54, 58-60, 64 and 65, and satisfy the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.

Preferably, the genomic CpG sequences are located within at least one gene sequence selected from the group consisting of *APC, CDKN2A, MYODI, CALCA, ESRI,*

35   *MGMT* and *TIMP3*, and combinations thereof. Preferably, the genomic CpG sequences located within at least one gene sequence selected from the group consisting of *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3*, correspond to genomic CpG sequences of CpG islands. Preferably, the *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3* gene sequences are those defined by the specific oligonucleotide

3

primers and probes corresponding to SEQ ID NOs:19-21, SEQ ID NOs:1-3, SEQ ID
NOs:7-9, SEQ ID NOs:10-12, SEQ ID NOs:4-6, SEQ ID NOs:16-18 and SEQ ID
NOs:13-15, respectively, as listed in TABLE II. Preferably, the CpG islands are located
within the promoter regions of the genes. Preferably, the *APC, CDKN2A, MYODI,*

5   *CALCA, ESRI, MGMT* and *TIMP3* gene sequences correspond to any CpG island
sequences associated with the sequences defined by the specific oligonucleotide primers
and probes corresponding to SEQ ID NOs:19-21, SEQ ID NOs:1-3, SEQ ID NOs:7-9,
SEQ ID NOs:10-12, SEQ ID NOs:4-6, SEQ ID NOs:16-18 and SEQ ID NOs:13-15,
respectively, as listed in TABLE II, or portions thereof, wherein the associated CpG island

10   sequences are those contiguous sequences of genomic DNA that encompass at least one
nucleotide of the sequences defined by the specific oligonucleotide primers and probes
corresponding to SEQ ID NOs:19-21, SEQ ID NOs:1-3, SEQ ID NOs:7-9, SEQ ID
NOs:10-12, SEQ ID NOs:4-6, SEQ ID NOs:16-18 and SEQ ID NOs:13-15, and satisfy the
criteria of having both a frequency of CpG dinucleotides corresponding to an

15   Observed/Expected Ratio >0.6, and a GC Content >0.5.

        Preferably, the cancer or cancer-related condition is selected from the group
consisting of gastrointestinal or esophageal adenocarcinoma, gastrointestinal or
esophageal dysplasia, gastrointestinal or esophageal metaplasia, Barrett's intestinal tissue,
pre-cancerous conditions in normal esophageal squamous mucosa, and combinations

20   thereof. Preferably, the cancer is esophageal adenocarcinoma, and wherein making a
diagnostic or prognostic prediction of the cancer, based upon the methylation state of the
genomic CpG sequences provides for classification of the adenocarcinoma by grade or
stage.

        Preferably, the methylation assay used to determine the methylation state of

25   genomic CpG sequences is selected from the group consisting of "MethylLight™", MS-
SNuPE, MSP, COBRA, MCA, and DMH, and combinations thereof.

        Preferably, the methylation assay used to determine the methylation state of
genomic CpG sequences is based, at least in part, on an array or microarray comprising
CpG sequences located within at least one gene sequence selected from the group

30   consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1,*
MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS and
MTHFR. Preferably, the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1,*
HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, and
TYMS gene sequences correspond to any CpG island sequences associated with the

35   sequences defined by the specific oligonucleotide primers and probes corresponding to
SEQ ID Nos:1-54, 58-60, 64 and 65, as listed in TABLE II, or portions thereof, wherein
the associated CpG island sequences are those contiguous sequences of genomic DNA that
encompass at least one nucleotide of the sequences defined by the specific oligonucleotide
primers and probes corresponding to SEQ ID Nos:1-54, 58-60, 64 and 65, and satisfy the

4

criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5. Preferably, the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR* gene sequences are

5   those defined by, or correspond to the specific oligonucleotide primers and probes corresponding to SEQ ID Nos:1-60, 64 and 65, as listed in TABLE II, or portions thereof.

Preferably, the methylation state of genomic CpG sequences that is determined is that of hypermethylation, hypomethylation or normal methylation.

The present invention also provides a kit useful for diagnosis or prognosis of

10   cancer or cancer-related conditions, comprising a carrier means containing one or more containers comprising: (a) a container containing a probe or primer which hybridizes to any region of a sequence located within at least one gene sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and

15   *MTHFR*; and (b) additional standard methylation assay reagents required to affect detection of methylated CpG-containing nucleic acid based, at least in part, on the probe or primer. Preferably, the additional standard methylation assay reagents are standard reagents for performing a methylation assay from the group consisting of MethyLight™, MS-SNuPE, MSP, COBRA, MCA and DMH, and combinations thereof. Preferably, the

20   probe or primer comprises at least about 12 to 15 nucleotides of a sequence selected from the group consisting of SEQ ID Nos:1-60, 64 and 65, as listed in TABLE II.

The present invention further provides a kit useful for diagnosis or prognosis of cancer or cancer-related conditions, comprising a carrier means containing one or more containers comprising: (a) an array or micorarray comprising sequences of at least about

25   12 to 15 nucleotides of a sequence selected from the group consisting of SEQ ID Nos:1-60, 64, 65, and any sequence located within a CpG island sequence associated with SEQ ID NOs:1-54, 58-60, 64 and 65.

BRIEF DESCRIPTION OF THE DRAWINGS

30   Figure 1 shows, according to the present invention, a quantitative methylation analysis of a panel of 20 genes from a screen of 84 tissue specimens from 31 patients with different stages of Barrett's esophagus ("IM"), dysplasia ("DYS") and/or associated esophageal adenocarcinoma ("T"). Methylation analysis was performed using the MethyLight™ assay (Eads et al., *Cancer Res.* 59:2302-2306, 1999; Eads et al., *Nucleic*

35   *Acids Res.* 28:E32, 2000). The percentage of fully methylated molecules at a specific locus (PMR = Percent of Methylated Reference) was calculated by dividing the *GENE/ACTB* ratio of a sample by the *GENE/ACTB* ratio of *SssI*-treated sperm DNA and multipling by 100. The resulting percentages were then dichotomized at 4% PMR to facilitate graphical representation and to reveal tissue-specific patterns (as described

herein). "N" indicates an analysis for which the control gene *ACTB* did not reach sufficient levels to allow the detection of a minimal value of 1 PMR for that methylation reaction in that particular sample.

Figure 2 shows the percent of samples methylated for each gene by tissue type.
5   The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. The genes, according to the present invention, were grouped according to their respective epigenetic gene classes (A-G) as shown in Figure 1. The letter "n" equals the number of samples analyzed for each tissue.

Figure 3 shows a comparison of epigenetic profiles according to the present
10   invention. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. Error bars represent the standard error of the mean. *Top panel*: Mean percent of genes methylated in each gene Class (A-F or ALL 19 CpG islands) by tissue type (N, normal esophagus; S, stomach; IM, intestinal metaplasia; DYS, dysplasia; T, adenocarcinoma). The error bars represent the standard
15   error of the mean (SEM). *Bottom panel*: Statistical analysis of the difference in mean percent of genes methylated in different tissues by gene Class (A-F) or for all 19 CpG islands combined (ALL). The *p*-values were generated by a Fisher's Protected Least Significant Difference (PLSD) test, adapted for use with unequal sample numbers (SAS Statview™ software).

20   Figure 4 shows the relationship between Class A methylation frequency and tumor stage according to the present invention. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. *Upper panel*: Mean number of genes methylated for Class A with respect to tumor stage (I-IV) is shown (see Figure 1). The error bars represent the standard error of the mean (SEM). The
25   letter "n" equals the number of samples analyzed in each tumor stage. *Lower panel*: Statistical analysis of the difference in mean number of Class A genes methylated by tumor stage. The *p*-values were generated by a Fisher's Protected Least Significant Difference (PLSD) test, adapted for use with unequal sample numbers (SAS Statview™ software).

30   Figure 5 shows, according to the present invention, the percent of two or more Class A genes methylated in intestinal metaplasia ("IM") tissues with ("Y"), or without ("N") associated dysplasia and/or adenocarcinoma. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. *Left panel*: Class A methylation in the IM data illustrated in Figure 1. Right panel: Class
35   A methylation in the IM for a completely independent follow-up study of twenty different microdissected IM samples. The error bars represent the standard error of the mean (SEM). The letter "n" equals the number of samples analyzed in each tissue group.

Figure 6 shows, according to the present invention, methylation frequency distributions in the progression of esophageal adenocarcinoma. The data was

6

dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4
PMR as unmethylated. The proportion of patients with zero to three (Class A), zero to nine
(Classes A + D) and zero to fourteen CpG islands (Classes A + B +C + D) methylated in
each tissue is shown. Class E and F CpG islands were not included since there was no
5   variation in the frequency of methylation between the different tissue. The letter "n"
equals the number of samples analyzed in each tissue.

DETAILED DESCRIPTION OF THE INVENTION

10

Definitions:
        The term "EAC" refers to esophageal adenocarcinoma, but also encompasses
different histological stages of esophageal adenocarcinoma corresponding to a multistep
process whereby normal squamous mucosa undergoes metaplasia to specialized columnar
15   epithelium (Intestinal Metaplasia (IM) or Barrett's esophagus), which then ultimately
progresses to dysplasia and subsequent malignancy (Barrett et al., *Nat. Genet.* 22:106-109,
1999; Zhuang et al., *Cancer Res.* 56:1961-4, 1996);
        The term "CIMP" refers to CpG island methylator phenotype, characterized by
widespread aberrant hypermethylation changes affecting multiple loci in a single tumor.
20   This is reflected in a bimodal distribution of the frequency of the number of genes
methylated in a group of tumors (16). CIMP tumors are a distinct group of tumors that are
defined by a high degree of concordant CpG island hypermethylation of genes exclusively
methylated in cancer, or type C genes. CIMP is now thought to be a new, distinct, yet
major pathway of tumorigenesis (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686,
25   1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999) (see "Background," above);
        The term "PMR" refers to percent of methylated reference, and is calculated as
described herein under Example I;
        "GC Content" refers, within a particular DNA sequence, to the [(number of C
bases + number of G bases) / band length for each fragment];
30   "Observed/Expected Ratio" ("O/E Ratio") refers to the frequency of CpG
dinucleotides within a particular DNA sequence, and corresponds to the [number of CpG
sites / (number of C bases X number of G bases)] X band length for each fragment;
        "CpG Island" refers to a contiguous region of genomic DNA that satisfies the
criteria of (1) having a frequency of CpG dinucleotides corresponding to an
35   "Observed/Expected Ratio" >0.6), and (2) having a "GC Content" >0.5. CpG islands are
typically, but not always, between about 0.2 to about 1 kb in length. A CpG island
sequence associated with a particular SEQ ID NO sequence of the present invention is that
contiguous sequence of genomic DNA that encompasses at least one nucleotide of the
particular SEQ ID NO sequence, and satisfies the criteria of having both a frequency of

7

CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6), and a GC Content >0.5;

"Methylation state" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence;

"Hypermethylation" refers to the methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample;

"Hypomethylation" refers to the methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample;

"Methylation assay" refers to any assay for determining the methylation state of a CpG dinucleotide within a sequence of DNA;

"MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997;

"MethyLight" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999;

"Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997;

"MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146;

"COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997;

"MCA" (Methylated CpG Island Amplification) refers to the methylation assay described by Toyota et al., *Cancer Res.* 59:2307-12, 1999, and in WO 00/26401A1;

"DMH" (Differential Methylation Hybridization) refers to the art-recognized methylation assay described in Huang et al., *Hum. Mol. Genet.*, 8:459-470, 1999, and in Yan et al., *Clin. Cancer Res.* 6:1432-38, 2000;

Genes and associated literature references:

"*APC*" refers to the adenomatous polyposis coli gene (Eads et al., *Cancer Res.* 59:2302-2306, 1999; Hiltunen et al., *Int. J. Cancer.* 70:644-648, 1997);

"*ARF*" refers to the P14 cell cycle regulator, tumor suppressor gene (Esteller et al., *Cancer Res.* 60:129-133, 2000; Robertson & Jones, *Mol. Cell. Biol.* 18:6457-6473, 1998);

8

"*CALCA*" refers to the calcitonin gene (Melki et al., *Cancer Res.* 59:3730-3740, 1999; Hakkarainen et al., *Int. J. Cancer.* 69:471-474, 1996);

"*CDH1*" refers to the E-cadherin gene (Melki et al., *Cancer Res.* 59:3730-3740, 1999; Ueki et al., *Cancer Res.* 60:1835-1839, 2000);

5   "*CDKN2A*" refers to the P16 gene (Jones & Laird, *Nat. Genet.* 21:163-167, 1999; Melki et al., *Cancer Res.* 59:3730-3740, 1999; Baylin & Herman, *Trends Genet.* 16:168-174, 2000; Cameron et al., *Nat. Genet.* 21:103-107, 1999; Ueki et al., *Cancer Res.* 60:1835-1839, 2000);

"*CDKN2B*" refers to the P15 gene (Melki et al., *Cancer Res.* 59:3730-3740, 1999; 10   Cameron et al., *Nat. Genet.* 21:103-107, 1999);

"*CTNNB1*" refers to the beta-catenin gene;

"*ESR1*" refers to the estrogen receptor alpha gene (Jones & Laird, *Nat. Genet.* 21:163-167, 1999; Baylin & Herman, *Trends Genet.* 16:168-174, 2000);

"*GSTP1*" refers to the glutathione S-transferase P1 gene (Melki et al., *Cancer Res.* 15   59:3730-3740, 1999; Tchou et al., *Int. J. Oncol.* 16:663-676, 2000);

"*HIC1*" refers to the hypermethylated in cancer 1 gene (Melki et al., *Cancer Res.* 59:3730-3740, 1999; Wales et al., *Nat. Med.* 1:570-577, 1995);

"*MGMT*" refers to the O6-methylguanine-DNA methyltransferase gene (Esteller et al., *Cancer Res.* 59:793-797, 1999);

20   "*MLH1*" refers to the Mut L homologue 1 gene (Jones & Laird, *Nat. Genet.* 21:163-167, 1999; Baylin & Herman, *Trends Genet.* 16:168-174, 2000; Cameron et al., *Nat. Genet.* 21:103-107, 1999; Esteller et al., *Am. J. Pathol.* 155:1767-1772, 1999, Ueki et al., *Cancer Res.* 60:1835-1839, 2000);

"*MTHFR*" refers to the methyl-tetrahydrofolate reductase gene (Pereira et al., 25   *Oncol. Rep.* 6:597-599, 1999);

"*MYOD1*" refers to the myogenic determinant 1 gene (Eads et al., *Cancer Res.* 59:2302-2306, 1999; Cheng et al., *Br. J. Cancer.* 75:396-402, 1997);

"*PTGS2*" refers to the cyclooxygenase 2 gene (Zimmermann et al., *Cancer Res.* 59:198-204, 1999);

30   "*RB1*" refers to the retinoblastoma gene (Stirzaker et al., *Cancer Res.* 57:2229-2237, 1997; Sakai et al., *Am. J. Hum. Genet.* 48:880-888, 1991);

"*TGFBR2*" refers to the transforming growth factor beta receptor II gene (Kang et al., *Oncogene.* 18:7280-7286, 1999; Hougaard et al., *Br. J. Cancer.* 79:1005-1011, 1999);

"*THBS1*" refers to the thrombospondin 1 gene (Ueki et al., *Cancer Res.* 60:1835-35   1839, 2000; Li et al., *Oncogene.* 18:284-3289, 1999);

"*TIMP3*" refers to the tissue inhibitor of metallinoproteinase 3 gene (Cameron et al., *Nat. Genet.* 21:103-107, 1999; Ueki et al., *Cancer Res.* 60:1835-1839, 2000; Bachman et al., *Cancer Res.* 59:798-802, 1999);

"*TYMS1*" refers to the thymidylate synthetase gene (Sakamoto et al., *In:* L. Herrera

(ed.) Familial adenomatous polyposis, pp. 315-324. New York: Alan R. Liss, 1990).

Overview

5   The present invention encompasses a broad, multi-gene approach that provides novel and therapeutically useful insight into concordant methylation behavior between and among genes. In particular embodiments, the present invention provides *novel epigenomic fingerprints* for the different histological stages of esophageal adenocarcinoma (EAC).

           More specifically, the present invention combines the advantages of both targeted

10   and comprehensive approaches by analyzing 20 different genes (*see* Table 1, below) using a quantitative, high-throughput methylation assay, "MethyLight™" (Eads et al., *Cancer Res.* 59:2302-2306, 1999; Eads et al., *Cancer Res.* 60:5021-5026, 2000; Eads et al., *Nucleic Acids Res.* 28:E32, 2000), to (i) more extensively characterize the methylation changes in esophageal adenocarcinoma (EAC); to (ii) generate epigenomic fingerprints for

15   the different histological stages of EAC; to (iii) identify epigenetic biomarkers useful in disease diagnosis and prevention; and to (iv) determine if CIMP is a contributor to the tumorigenesis of esophageal adenocarcinoma tumors.

           A total of 104 tissue specimens from 51 patients with different stages of Barrett's esophagus and/or associated adenocarcinoma were analyzed. Specifically, 84 of these

20   tissue specimens were screened with the full panel of 20 genes, revealing distinct classes of methylation patterns in the different types of tissue.

           The most informative genes, for purposes of the present invention, were those with an intermediate frequency of significant hypermethylation (*i.e.*, those ranging from about 15% (*CDKN2A*) to about 60% (*MGMT*) of the samples). This group of genes could be

25   further subdivided into three classes, according to the (1) absence (*CDKN2A, ESR1* and *MYOD1*), or (2) presence (*CALCA, MGMT* and *TIMP3*) of methylation in normal esophageal mucosa and stomach, or (3) the infrequent methylation of normal esophageal mucosa accompanied by methylation in all normal stomach samples (*APC*).

           The other genes were relatively less informative, since the frequency of

30   hypermethylation was below about 5% (*ARF, CDH1, CDKN2B, GSTP1, MLH1, PTGS2* and *THBS1*), completely absent (*CTNNB1, RB1, TGFBR2* and *TYMS1*) or ubiquitous (*HIC1* and *MTHFR*), regardless of tissue type.

           Each class of gene undergoes unique epigenetic changes at different steps of disease progression of EAC, consistent with a step-wise loss of multiple protective barriers

35   against CpG island hypermethylation. The aberrant hypermethylation occurs at many different loci in the same tissues, consistent with an overall deregulation of methylation control in EAC tumorigenesis. However, there was no clear evidence for a distinct group of tumors with a CpG island methylator phenotype ("CIMP").

           Additionally, normal and metaplastic tissues from patients with evidence of

10

associated dysplasia or cancer displayed a significantly higher incidence of hypermethylation than similar tissues from patients with no further progression of their disease. The fact that the samples from these two groups of patients were histologically indistinguishable, yet molecularly distinct, indicates, according to the present invention,

5   that the occurrence of such hypermethylation provides a novel and valuable clinical tool to identify patients with pre-malignant Barrett's, who are at risk for further progression.

       TABLE I shows a list of gene names and functions analyzed by the MethyLight™ assay in EAC. The genes are listed in alphabetical order based on their designated HUGO (HUman Genome Organization) names. The genes are divided into three groups

10   according to whether or not they have CpG islands and are known to be methylated in other tumors. A brief description of the function of each gene is included.

TABLE I. List of Gene Names Analyzed by the MethyLight™ assay in EAC

GENE SYMBOL	NAME	TUMORS WITH HYPERMETHYLATION †	PROPOSED FUNCTION
CPG ISLAND: METHYLATED IN TUMORS			
APC	Adenomatous Polyposis Coli	Colon	WNT Signaling Pathway; Beta-Catenin Degradation; Tumor Suppressor
ARF	P14	Colon, Lymphoma	Cell Cycle Regulator; Tumor Suppressor
CALCA	Calcitonin	Breast, Colon, Leukemia	Regulates Calcium Levels via Adenylate Cyclase
CDH1	E-Cadherin	AML, Bladder, Breast, Colon, Gastric, Thyroid	Cell Adhesion
CDKN2A	P16	AML, Bladder, Colon, Gastric, Lymphoma, Melanoma	Cell Cycle Regulator, Tumor Suppressor
CDKN2B	P15	Colon, Hematological Malignancies	Cell Cycle Regulator, Tumor Suppressor
ESR1	Estrogen Receptor Alpha	AML, Colon, Breast, Lung, Leukemia	Hormone Receptor in Mammary Cells; Growth Suppressive in Colorectal Cancer Cells
GSTP1	Glutathione S-Transferase PI	Breast, Prostate, Hepatocellular	Protects Against Oxidant and Electrophilic Carcinogens
HIC1	Hypermethylated in Cancer 1	Brain, Breast, Colon, Renal, Leukemia, Lymphoma	Zinc Finger Transcription Factor; Potential Tumor Suppressor
MGMT	O6-Methylguanine-DNA Methyltransferase	Brain, Colon, Lymphoma, Non-Small Cell Lung Cancer	DNA Repair
MLH1	Mut L Homologue 1	Colon, Endometrial, Gastric	DNA Mismatch Repair
MYOD1	Myogenic Determinant 1	Breast, Colon, Ovarian	Muscle Differentiation
RB1	Retinoblastoma 1	Retinoblastoma	Cell Cycle Regulator; Tumor Suppressor

GENE SYMBOL	NAME	TUMORS WITH HYPERMETHYLATION [†]	PROPOSED FUNCTION
TGFBR2	Transforming Growth Factor Beta Receptor II	Colon, Gastric, Small-Cell Lung Cancer	Serine-Threonine Kinase Receptor; Cell Signaling
THBS1	Thrombospondin 1	Colon, Glioblastomas	Angiogenesis Inhibitor
TIMP3	Tissue Inhibitor of Metallinoproteinase 3	Brain, Breast, Colon, Kidney, Lung, Pancreatic	Metastasis, Angiogenesis
CPG ISLAND: UNKNOWN METHYLATION STATUS			
CTNNB1	Beta-Catenin	Unknown	WNT Signaling Pathway (Regulated by APC); Cell Adhesion via E-Cadherin
PTGS2	Cyclooxygenase 2	Unknown	Prostaglandin Synthesis from Arachidonic Acid
TYMS	Thymidylate Synthetase	Unknown	Nucleotide/DNA Synthesis
NON-CPG ISLAND: UNKNOWN METHYLATION STATUS			
MTHFR	Methyl-Tetetrahydrofolate Reductase	Unknown	Folate Metabolism, Methionine Synthesis; Predisposing Polymorphism in Colon Cancer

TABLE I cont.

† *See literature references relating to specific genes under "DEFINITIONS," herein above.*

13

Diagnostic and Prognostic Assays for Cancer

   The present invention provides for diagnostic and prognostic cancer assays based on determination of the methylation state of one or more of the disclosed 20 gene sequences
5 (*APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR; see* TABLES I and II, below; *and see* under "Definitions," above), or methylation-altered DNA sequence embodiments thereof. These 20 gene sequence regions are defined herein by the oligomeric primers and probes corresponding to SEQ ID NOS:1-60, 64 and 65 (*see* TABLE
10 II, below). SEQ ID NOS:61-63 correspond to the *ACTB* "control" gene region used in the present analysis (*see* EXAMPLE 1, below).
   Additionally, 19 of these 20 gene sequence regions correspond to CpG islands or regions thereof (based on GC Content and O/E ratio); namely *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1,*
15 *TIMP3, CTNNB1, PTGS2* and *TYMS* (*see* TABLE 1, below). Thus, based on the fact that the methylation state of a portion of a given CpG island is generally representative of the island as a whole, the present invention further encompasses the novel use of *any* sequences within the 19 complete CpG islands associated with these 19 gene sequence regions (defined herein by the primers and probes corresponding to SEQ ID NOS:1-60, 64 and 65 (*see* TABLE II,
20 below) in cancer prognostic and diagnostic applications), where a CpG island sequence associated with one of these 19 gene sequences is that contiguous sequence of genomic DNA that encompasses at least one nucleotide of one of these 19 gene sequences, and satisfies the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.
25   Typically, such assays involve obtaining a tissue sample from a test tissue, performing a methylation assay on DNA derived from the tissue sample to determine the associated methylation state, and making a diagnosis or prognosis based thereon.
   The methylation assay is used to determine the methylation state of one or a plurality of CpG dinucleotide within a DNA sequence of the DNA sample. According to the present
30 invention, possible methylation states include *hypermethylation* and *hypomethylation*, relative to a normal state (*i.e.,* non-cancerous control state). Hypermethylation and hypomethylation refer to the methylation states corresponding to an *increased* or *decreased*, respectively, presence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence of the test sample, relative to the amount of 5-mCyt found at corresponding
35 CpG dinucleotides within a normal control DNA sample.
   A diagnosis or prognosis is based, at least in part, upon the determined methylation

14

state of the sample DNA sequence compared to control data obtained from normal, non-cancerous tissue.

Methylation Assay Procedures

5

Various methylation assay procedures are known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides within a DNA sequence (*e.g.*, CpG islands). Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, use of methylation-sensitive restriction enzymes, etc.

10

For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, *e.g.*, the method described by Sadri & Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

15

Preferably, assays such as "MethyLight^{TM}" (a fluorescence-based real-time PCR technique) (Eads et al., *Cancer Res.* 59:2302-2306, 1999), Methylation-sensitive Single Nucleotide Primer Extension reactions ("Ms-SnuPE"; Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997), methylation-specific PCR ("MSP"; Herman et al., *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146), and methylated CpG island amplification ("MCA";Toyota et al., *Cancer Res.* 59:2307-12, 1999) are used alone or in combination with other of these methods. Methylation assays that can be used in various embodiments of the present invention include, but are not limited to, the following assays.

20

25

COBRA (Combined Bisulfite Restriction Analysis): COBRA analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the interested CpG islands, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of

30

35

15

digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum
of DNA methylation levels. Additioinally, this technique can be reliably applied to DNA
obtained from microdissected paraffin-embedded tissue samples. Typical reagents (*e.g.*, as
might be found in a typical COBRA-based methylation kit) for COBRA analysis may include,

5   but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence
or CpG island); restriction enzyme and appropriate buffer; gene-hybridization oligo; control
hybridization oligo; kinase labeling kit for oligo probe; and radioactive nucleotides (although
other label schemes known in the art including, but not limited, to fluorescent and
phosphorescent schemes can be used). Additionally, bisulfite conversion reagents may

10   include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (*e.g.*,
precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery
components.

   Ms-SnuPE (Methylation-sensitive Single Nucleotide Primer Extension). The Ms-
SNuPE technique is a quantitative method for assessing methylation differences at specific

15   CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer
extension (Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997). Briefly, genomic
DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving
5-methylcytosine unchanged. Amplification of the desired target sequence is then performed
using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated

20   and used as a template for methylation analysis at the CpG site(s) of interest. Small amounts
of DNA can be analyzed (*e.g.*, microdissected pathology sections), and it avoids utilization of
restriction enzymes for determining the methylation status at CpG sites. Typical reagents
(*e.g.*, as might be found in a typical Ms-SNuPE-based methylation kit) for Ms-SNuPE
analysis may include, but are not limited to: PCR primers for specific gene (or methylation-

25   altered DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel
extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer
(for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion
reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or
kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA

30   recovery components.

   MSP (Methylation-specific PCR). MSP allows for assessing the methylation status of
virtually any group of CpG sites within a CpG island, independent of the use of methylation-
sensitive restriction enzymes (Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996;
US Patent No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all

35   unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers
specific for methylated versus unmethylated DNA. MSP requires only small quantities of

16

DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples. Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA

5   sequence or CpG island), optimized PCR buffers and deoxynucleotides, and specific probes.

MCA (Methylated CpG Island Amplification). The MCA technique is a method that can be used to screen for altered methylation patterns in genomic DNA, and to isolate specific sequences associated with these changes (Toyota et al., *Cancer Res.* 59:2307-12, 1999). Briefly, restriction enzymes with different sensitivities to cytosine methylation in their

10   recognition sites are used to digest genomic DNAs from primary tumors, cell lines, and normal tissues prior to arbitrarily primed PCR amplification. Fragments that show differential methylation are cloned and sequenced after resolving the PCR products on high-resolution polyacrylamide gels. The cloned fragments are then used as probes for Southern analysis to confirm differential methylation of these regions. Typical reagents (*e.g.*, as might be found in

15   a typical MCA-based kit) for MCA analysis may include, but are not limited to: PCR primers for arbitrary priming Genomic DNA; PCR buffers and nucleotides, restriction enzymes and appropriate buffers; gene-hybridization oligos or probes; control hybridization oligos or probes.

DMH (Differential Methylation Hybridization). DMH refers to the art-recognized,

20   array-based methylation assay described in Huang et al., *Hum. Mol. Genet.*, 8:459-470, 1999, and in Yan et al., *Clin. Cancer Res.* 6:1432-38, 2000. DMH allows for a genome-wide screening of CpG island hypermethylation in cancer cell lines, and. Briefly, CpG island tags are arrayed on solid supports (*e.g.*, nylon membranes, silicon, etc.), and probed with "amplicons" representing a pool of methylated CpG DNA, from test (*e.g.*, tumor) or reference

25   samples. The differences in test and reference signal intensities on screened CpG island arrays reflect methylation alterations of corresponding sequences in the test DNA.

MethyLight^{TM}. In preferred embodiments, the MethyLight^{TM} assay is used to determine the methylation status of one or more CpG sequences. The MethyLight^{TM} assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time

30   PCR (TaqMan ®) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 60:5021-5026, 2000; Eads et al., *Cancer Res.* 59:2302-2306, 1999; Eads et al., *Nucleic Acids Res.* 28:E32, 2000). Briefly, the MethyLight^{TM} process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the

35   bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an "unbiased" (with primers that do not overlap known CpG

17

methylation sites) PCR reaction, or in a "biased" (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

     The MethyLight™ assay may assay be used as a quantitative test for methylation
5   patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a
10   qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not "cover" known methylation sites (a fluorescence-based version of the "MSP" technique), or with oligonucleotides covering potential methylation sites.

     The MethyLight™ process can by used with a "TaqMan®" probe in the amplification
15   process. For example, double-stranded genomic DNA is treated with sodium bisulfite and subjected to one of two sets of PCR reactions using TaqMan® probes; e.g., with either biased primers and TaqMan® probe, or unbiased primers and TaqMan® probe. The TaqMan® probe is dual-labeled with fluorescent "reporter" and "quencher" molecules, and is designed to be specific for a relatively high GC content region so that it melts out at about 10°C higher
20   temperature in the PCR cycle than the forward or reverse primers. This allows the TaqMan® probe to remain fully hybridized during the PCR annealing/extension step. As the Taq polymerase enzymatically synthesizes a new strand during PCR, it will eventually reach the annealed TaqMan® probe. The Taq polymerase 5' to 3' endonuclease activity will then displace the TaqMan® probe by digesting it to release the fluorescent reporter molecule for
25   quantitative detection of its now unquenched signal using a real-time fluorescent detection system.

     Typical reagents (e.g., as might be found in a typical MethyLight™ -based methylation kit) for MethyLight™ analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes;
30   optimized PCR buffers and deoxynucleotides; and Taq polymerase. A detailed description of four alternate process applications ("A" through "D") of the MethyLight™ assay follows below. Preferably, the quantitative MethyLight™ process application "B" is used.

     MethyLight™-based detection of the methylated nucleic acid is relatively rapid and is based on amplification-mediated displacement of specific oligonucleotide probes. In a
35   preferred embodiment, amplification and detection, in fact, occur simultaneously as measured by fluorescence-based real-time quantitative PCR ("RT-PCR") using specific, dual-labeled

TaqMan® oligonucleotide probes, with no requirement for subsequent manipulation or analysis. The displaceable probes can be specifically designed to distinguish between methylated and unmethylated CpG sites present in the original, unmodified nucleic acid sample.

5   Like the technique of methylation-specific PCR ("MSP"; US Patent 5,786,146), MethyLight™ provides for significant advantages over previous PCR-based and other methods (e.g., Southern analyses) used for determining methylation patterns. MethyLight™ is substantially more sensitive than Southern analysis, and facilitates the detection of a low number (percentage) of methylated alleles in very small nucleic acid samples, as well as

10   paraffin-embedded samples. Moreover, in the case of genomic DNA, analysis is not limited to DNA sequences recognized by methylation-sensitive restriction endonucleases, thus allowing for fine mapping of methylation patterns across broader CpG-rich regions. MethyLight™ also eliminates any false-positive results, that otherwise might result from incomplete digestion by methylation-sensitive restriction enzymes, inherent in previous PCR-

15   based methylation methods.

MethyLight™ can be applied as a quantitative process for measuring methylation amounts, and is substantially more rapid than other methods. MethyLight™ does not require any post-PCR manipulation or processing. This not only greatly reduces the amount of labor involved in the analysis of bisulfite-treated DNA, but it also provides a means to avoid

20   handling of PCR products that could contaminate future reactions.

One process embodiment uses MethyLight™ for the unbiased amplification of all possible methylation states using primers that do not cover any CpG sequences in the original, unmodified DNA sequence. To the extent that all methylation patterns are amplified equally, quantitative information about DNA methylation patterns are then distilled from the resulting

25   PCR pool by any technique capable of detecting sequence differences (e.g., by fluorescence-based PCR).

MethyLight™ employs one or a series of CpG-specific TaqMan® probes, each corresponding to a particular methylation site in a given amplified DNA region, are constructed. This series of probes is then utilized in parallel amplification reactions, using

30   aliquots of a single, modified DNA sample, to simultaneously determine the complete methylation pattern present in the original unmodified sample of genomic DNA. This is accomplished in a fraction of the time and expense required for direct sequencing of the sample of genomic DNA, and are substantially more sensitive. Moreover, one embodiment of MethyLight™ provides for a quantitative assessment of such a methylation pattern.

35   The present invention, as described herein, may be practiced using a variety of methylation assays. For MethyLight™ emabodiments, there are four process techniques and

19

associated diagnostic kits that a methylation-dependent nucleic acid modifying agent (*e.g.*, bisulfite), to both qualitatively and quantitatively determine CpG methylation status in nucleic acid samples (*e.g.*, genomic DNA samples). The four processes are described herein as processes "A," "B," "C" and "D." Overall, methylated-CpG sequence discrimination is

5   designed to occur at the level of amplification, probe hybridization or at both levels. For example, applications C and D utilize "biased" primers that distinguish between modified unmethylated and methylated nucleic acid and provide methylated-CpG sequence discrimination at the PCR amplification level. Process B uses "unbiased" primers (that do not cover CpG methylation sites), to provide for unbiased amplification of modified nucleic acid,

10   but rather utilize probes that distinguish between modified unmethylated and methylated nucleic acid to provide for quantitative methylated-CpG sequence discrimination at the detection level (*e.g.*, at the fluorescent (or luminescent) probe hybridization level only). Process A does not, in itself, provide for methylated-CpG sequence discrimination at either the amplification or detection levels, but supports and validates the other three applications by

15   providing control reactions for input DNA.

MethyLight[TM] *Process D.* In a first MethyLight[TM] embodiment, the invention provides a method for qualitatively detecting a methylated CpG-containing nucleic acid, the method including: contacting a nucleic acid-containing sample with a modifying agent that modifies unmethylated cytosine to produce a converted nucleic acid; amplifying the converted nucleic

20   acid by means of two oligonucleotide primers in the presence of a specific oligonucleotide hybridization probe, wherein both the primers and probe distinguish between modified unmethylated and methylated nucleic acid; and detecting the "methylated" nucleic acid based on amplification-mediated probe displacement.

The term "modifies" as used herein means the conversion of an unmethylated cytosine

25   to another nucleotide by the modifying agent, said conversion distinguishing unmethylated from methylated cytosine in the original nucleic acid sample. Preferably, the agent modifies unmethylated cytosine to uracil. Preferably, the agent used for modifying unmethylated cytosine is sodium bisulfite, however, other equivalent modifying agents that selectively modify unmethylated cytosine, but not methylated cytosine, can be substituted in the method

30   of the invention. Sodium-bisulfite readily reacts with the 5, 6-double bond of cytosine, but not with methylated cytosine, to produce a sulfonated cytosine intermediate that undergoes deamination under alkaline conditions to produce uracil. Because Taq polymerase recognizes uracil as thymine and 5-methylcytidine (^{m5}C) as cytidine, the sequential combination of sodium bisulfite treatment and PCR amplification results in the ultimate conversion of

35   unmethylated cytosine residues to thymine (C $\rightarrow$ U $\rightarrow$ T) and methylated cytosine residues ("^{m}C") to cytosine (^{m}C $\rightarrow$ ^{m}C $\rightarrow$ C). Thus, sodium-bisulfite treatment of genomic DNA

20

creates methylation-dependent sequence differences by converting unmethylated cyotsines to uracil, and upon PCR the resultant product contains cytosine only at positions where methylated cytosine occurs in the unmodified nucleic acid.

Oligonucleotide "primers," as used herein, means linear, single-stranded, oligomeric

5   deoxyribonucleic or ribonucleic acid molecules capable of sequence-specific hybridization (annealing) with complementary strands of modified or unmodified nucleic acid. As used herein, the specific primers are preferably DNA. The primers of the invention embrace oligonucleotides of appropriate sequence and sufficient length so as to provide for specific and efficient initiation of polymerization (primer extension) during the amplification process.

10 As used in the inventive processes, oligonucleotide primers typically contain 12-30 nucleotides or more, although may contain fewer nucleotides. Preferably, the primers contain from 18-30 nucleotides. The exact length will depend on multiple factors including temperature (during amplification), buffer, and nucleotide composition. Preferably, primers are single-stranded although double-stranded primers may be used if the strands are first

15 separated. Primers may be prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments which are commonly known in the art.

As used in the inventive embodiments herein, the specific primers are preferably designed to be substantially complementary to each strand of the genomic locus of interest.

20 Typically, one primer is complementary to the negative (-) strand of the locus (the "lower" strand of a horizontally situated double-stranded DNA molecule) and the other is complementary to the positve (+) strand ("upper" strand). As used in the embodiment of Application D, the primers are preferably designed to overlap potential sites of DNA methylation (CpG nucleotides) and specifically distinguish modified unmethylated from

25 methylated DNA. Preferably, this sequence discrimination is based upon the differential annealing temperatures of perfectly matched, versus mismatched oligonucleotides. In the embodiment of Application D, primers are typically designed to overlap from one to several CpG sequences. Preferably, they are designed to overlap from 1 to 5 CpG sequences, and most preferably from 1 to 4 CpG sequences. By contrast, in a quantitative embodiment of the

30 invention employed in the Examples of the present invention, the primers do not overlap any CpG sequences.

In the case of fully "unmethylated" (complementary to modified unmethylated nucleic acid strands) primer sets, the anti-sense primers contain adenosine residues ("A"s) in place of guanosine residues ("G"s) in the corresponding (-) strand sequence. These substituted As in

35 the anti-sense primer will be complementary to the uracil and thymidine residues ("Us" and "Ts") in the corresponding (+) strand region resulting from bisulfite modification of

unmethylated C residues ("Cs") and subsequent amplification. The sense primers, in this
case, are preferably designed to be complementary to anti-sense primer extension products,
and contain Ts in place of unmethylated Cs in the corresponding (+) strand sequence. These
substituted Ts in the sense primer will be complementary to the As, incorporated in the anti-
5   sense primer extension products at positions complementary to modified Cs (Us) in the
original (+) strand.

In the case of fully-methylated primers (complementary to methylated CpG-containing
nucleic acid strands), the anti-sense primers will not contain As in place of Gs in the
corresponding (-) strand sequence that are complementary to methylated Cs (*i.e.*, ^{m}CpG
10   sequences) in the original (+) strand. Similarly, the sense primers in this case will not contain
Ts in place of methylated Cs in the corresponding (+) strand ^{m}CpG sequences. However, Cs
that are not in CpG sequences in regions covered by the fully-methylated primers, and are not
methylated, will be represented in the fully-methylated primer set as described above for
unmethylated primers.

15   Preferably, as employed in the embodiment of process D, the amplification process
provides for amplifying bisulfite converted nucleic acid by means of two oligonucleotide
primers in the presence of a specific oligonucleotide hybridization probe. Both the primers
and probe distinguish between modified unmethylated and methylated nucleic acid.
Moreover, detecting the "methylated" nucleic acid is based upon amplification-mediated
20   probe fluorescence. In one embodiment, the fluorescence is generated by probe degradation
by 5' to 3' exonuclease activity of the polymerase enzyme. In another embodiment, the
fluorescence is generated by fluorescence energy transfer effects between two adjacent
hybridizing probes (Lightcycler® technology) or between a hybridizing probe and a primer.
In another embodiment, the fluorescence is generated by the primer itself (Sunrise®
25   technology). Preferably, the amplification process is an enzymatic chain reaction that uses the
oligonucleotide primers to produce exponential quantities of amplification product, from a
target locus, relative to the number of reaction steps involved.

As describe above, one member of a primer set is complementary to the (-) strand,
while the other is complementary to the (+) strand. The primers are chosen to bracket the area
30   of interest to be amplified; that is, the "amplicon." Hybridization of the primers to denatured
target nucleic acid followed by primer extension with a DNA polymerase and nucleotides,
results in synthesis of new nucleic acid strands corresponding to the amplicon. Preferably, the
DNA polymerase is *Taq* polymerase, as commonly used in the art. Although equivalent
polymerases with a 5' to 3' nuclease activity can be substituted. Because the new amplicon
35   sequences are also templates for the primers and polymerase, repeated cycles of denaturing,
primer annealing, and extension results in exponential production of the amplicon. The

22

product of the chain reaction is a discrete nucleic acid duplex, corresponding to the amplicon sequence, with termini defined by the ends of the specific primers employed. Preferably the amplification method used is that of PCR (Mullis et al., *Cold Spring Harb. Symp. Quant. Biol.* 51:263-273; Gibbs, *Anal. Chem.* 62:1202-1214, 1990), or more preferably, automated

5   embodiments thereof which are commonly known in the art.

Preferably, methylation-dependent sequence differences are detected by methods based on fluorescence-based quantitative PCR (real-time quantitative PCR, Heid et al., *Genome Res.* 6:986-994, 1996; Gibson et al., *Genome Res.* 6:995-1001, 1996) (*e.g.*, "TaqMan®," "Lightcycler®," and "Sunrise®" technologies). For the TaqMan® and

10   Lightcycler® technologies, the sequence discrimination can occur at either or both of two steps: (1) the amplification step, or (2) the fluorescence detection step. In the case of the "Sunrise®" technology, the amplification and fluorescent steps are the same. In the case of the FRET hybridization, probes format on the Lightcycler®, either or both of the FRET oligonucleotides can be used to distinguish the sequence difference. Most preferably the

15   amplification process, as employed in all inventive embodiments herein, is that of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996) employing a dual-labeled fluorescent oligonucleotide probe (TaqMan® PCR, using an ABI Prism 7700 Sequence Detection System, Perkin Elmer Applied Biosystems, Foster City, California).

20   The "TaqMan®" PCR reaction uses a pair of amplification primers along with a nonextendible interrogating oligonucleotide, called a TaqMan® probe, that is designed to hybridize to a GC-rich sequence located between the forward and reverse (*i.e.*, sense and anti-sense) primers. The TaqMan® probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (*e.g.*, phosphoramidites) attached to

25   nucleotides of the TaqMan® oligonucleotide. Examples of suitable reporter and quencher molecules are: the 5' fluorescent reporter dyes 6FAM ("FAM"; 2,7 dimethoxy-4,5-dichloro-6-carboxy-fluorescein), and TET (6-carboxy-4,7,2',7'-tetrachlorofluorescein); and the 3' quencher dye TAMRA (6-carboxytetramethylrhodamine) (Livak et al., *PCR Methods Appl.* 4:357-362, 1995; Gibson et al., *Genome Res.* 6:995-1001; and 1996; Heid et al., *Genome Res.*

30   6:986-994, 1996).

One process for designing appropriate TaqMan® probes involves utilizing a software facilitating tool, such as "Primer Express" that can determine the variables of CpG island location within GC-rich sequences to provide for at least a 10°C melting temperature difference (relative to the primer melting temperatures) due to either specific sequence

35   (tighter bonding of GC, relative to AT base pairs), or to primer length.

The TaqMan® probe may or may not cover known CpG methylation sites, depending

on the particular inventive process used. Preferably, in the embodiment of process D, the TaqMan® probe is designed to distinguish between modified unmethylated and methylated nucleic acid by overlapping from 1 to 5 CpG sequences. As described above for the fully unmethylated and fully methylated primer sets, TaqMan® probes may be designed to be

5   complementary to either unmodified nucleic acid, or, by appropriate base substitutions, to bisulfite-modified sequences that were either fully unmethylated or fully methylated in the original, unmodified nucleic acid sample.

Each oligonucleotide primer or probe in the TaqMan® PCR reaction can span anywhere from zero to many different CpG dinucleotides that each can result in two different

10   sequence variations following bisulfite treatment (^{m}CpG, or UpG). For instance, if an oligonucleotide spans 3 CpG dinucleotides, then the number of possible sequence variants arising in the genomic DNA is $2^3 = 8$ different sequences. If the forward and reverse primer each span 3 CpGs and the probe oligonucleotide (or both oligonucleotides together in the case of the FRET format) spans another 3, then the total number of sequence permutations

15   becomes $8 \times 8 \times 8 = 512$. In theory, one could design separate PCR reactions to quantitatively analyze the relative amounts of each of these 512 sequence variants. In practice, a substantial amount of qualitative methylation information can be derived from the analysis of a much smaller number of sequence variants. Thus, in its most simple form, the inventive process can be performed by designing reactions for the fully methylated and the

20   fully unmethylated variants that represent the most extreme sequence variants in a hypothetical example. The ratio between these two reactions, or alternatively the ratio between the methylated reaction and a control reaction (process A), would provide a measure for the level of DNA methylation at this locus.

Detection of methylation in the MethyLight™ embodiment of process D, as in other

25   MethyLight™ embodiments herein, is based on amplification-mediated displacement of the probe. In theory, the process of probe displacement might be designed to leave the probe intact, or to result in probe digestion. Preferably, as used herein, displacement of the probe occurs by digestion of the probe during amplification. During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5' to 3' nucleolytic activity of the

30   DNA polymerase. On cleavage of the probe, the reporter moiety emission is no longer transferred efficiently to the quenching moiety, resulting in an increase of the reporter moiety fluorescent-emission spectrum at 518 nm. The fluorescent intensity of the quenching moiety (e.g., TAMRA), changes very little over the course of the PCR amplification. Several factors my influence the efficiency of TaqMan® PCR reactions including: magnesium and salt

35   concentrations; reaction conditions (time and temperature); primer sequences; and PCR target size (i.e., amplicon size) and composition. Optimization of these factors to produce the

24

optimum fluorescence intensity for a given genomic locus is obvious to one skilled in the art of PCR, and preferred conditions are further illustrated in the "Examples" herein. The amplicon may range in size from 50 to 8,000 base pairs, or larger, but may be smaller. Typically, the amplicon is from 100 to 1000 base pairs, and preferably is from 100 to 500

5   base pairs. Preferably, the reactions are monitored in real time by performing PCR amplification using 96-well optical trays and caps, and using a sequence detector (ABI Prism) to allow measurement of the fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Preferably, process D is run in combination with the process A to provide controls for the amount of input nucleic acid, and to normalize data

10   from tray to tray.

MethyLight^{TM} *Process C.* The MethyLight^{TM} process can be modified to avoid sequence discrimination at the PCR product detection level. Thus, in an additional qualitative process embodiment, just the primers are designed to cover CpG dinucleotides, and sequence discrimination occurs solely at the level of amplification. Preferably, the probe used in this

15   embodiment is still a TaqMan® probe, but is designed so as not to overlap any CpG sequences present in the original, unmodified nucleic acid. The embodiment of process C represents a high-throughput, fluorescence-based real-time version of MSP technology, wherein a substantial improvement has been attained by reducing the time required for detection of methylated CpG sequences. Preferably, the reactions are monitored in real time

20   by performing PCR amplification using 96-well optical trays and caps, and using a sequence detector (ABI Prism) to allow measurement of the fluorescent spectra of all 96 wells of the thermal cylcer continuously during the PCR amplification. Preferably, process C is run in combination with process A (below) to provide controls for the amount of input nucleic acid, and to normalize data from tray to tray.

25   *MethyLight*^{TM} *Process B.* In preferred embodiments of the present invention, the MethyLight^{TM} process can be also be modified to avoid sequence discrimination at the PCR amplification level. In a quantitative process B embodiment, just the probe is designed to cover CpG dinucleotides, and sequence discrimination occurs solely at the level of probe hybridization. Preferably, TaqMan® probes are used. In this version, sequence variants

30   resulting from the bisulfite conversion step are amplified with equal efficiency; as long as there is no inherent amplification bias (Warnecke et al., *Nucleic Acids Res.* 25:4422-4426, 1997). Design of separate probes for each of the different sequence variants associated with a particular methylation pattern (*e.g.*, $2^3=8$ probes in the case of 3 CpGs) would allow a quantitative determination of the relative prevalence of each sequence permutation in the

35   mixed pool of PCR products. Preferably, the reactions are monitored in real time by performing PCR amplification using 96-well optical trays and caps, and using a sequence

detector (ABI Prism) to allow measurement of the fluorescent spectra of all 96 wells of the thermal cylcer continuously during the PCR amplification. Preferably, process B is run in combination with process A, below to provide controls for the amount of input nucleic acid, and to normalize data from tray to tray.

5   *MethyLight^TM Process A*. MethyLight^TM process A does not, in itself, provide for methylated-CpG sequence discrimination at either the amplification or detection levels, but supports and validates the other three process applications by providing control reactions for the amount of input DNA, and to normalize data from tray to tray. Thus, if neither the primers, nor the probe overlie any CpG dinucleotides, then the reaction represents unbiased

10   amplification and measurement of amplification using fluorescent-based quantitative real-time PCR serves as a control for the amount of input DNA. Preferably, process A not only lacks CpG dinucleotides in the primers and probe(s), but also does not contain any CpGs within the amplicon at all to avoid any differential effects of the bisulfite treatment on the amplification process. Preferably, the amplicon for process A is a region of DNA that is not

15   frequently subject to copy number alterations, such as gene amplification or deletion.

Results obtained with the qualitative MethyLight^TM version (process embodiment "B" of the technology) are described in the Examples below. Dozens of human tumor samples have been analyzed using this technology with excellent results.

20

Cancer Diagnostic and Prognostic Assays and Kits

Typically, diagnostic and/or prognostic assays of the present invention involve obtaining a tissue sample from a test tissue, performing a methylation assay on DNA derived

25   from the tissue sample to determine the associated methylation state, and making a diagnosis or prognosis based thereon.

In preferred embodiments, diagnostic and prognostic cancer assays are based on determination of the methylation state of one or more of the disclosed 20 gene sequences (*APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1,*

30   *MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR,* or methylation-altered DNA sequence embodiments thereof), as defined herein by the oligomeric primers and probes corresponding to SEQ ID NOS:1-60, 64 and 65 (*see* TABLE II, below). SEQ ID NOS:61-63 correspond to the *ACTB* "control" gene region used in the present analysis (*see* EXAMPLE 1, below).

35   Additionally, other primers or probes corresponding to other sequence regions of the CpG islands associated with the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1,*

26

GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2
and *TYMS* sequence regions used herein may be used, based on the fact that the methylation
state of a portion of a given CpG island is generally representative of the island as a whole.

5   Accordingly, the reagents required to perform one or more art-recognized methylation
assays (including those described above) are combined with such primers and/or probes, or
portions thereof, to determine the methylation state of CpG-containing nucleic acids.

For example, the MethyLight™, Ms-SNuPE, MCA, COBRA, and MSP methylation
assays could be used alone or in combination, along with primers or probes comprising the
sequences of SEQ ID NOS:1-65, or portions thereof, to determine the methylation state of a
10   CpG dinucleotide within one or more of the 20 gene sequence regions corresponding to *APC,
ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1,
RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* or *MTHFR*, or, in the case of 19 of
these 20 sequence regions (*i.e.*, for all but MTHFR), to other CpG island sequences associated
with these sequences, where such other CpG island sequences associated with these 19 gene
15   sequences are those contiguous sequences of genomic DNA that encompasses at least one
nucleotide of one of these 19 gene sequence regions, and satisfy the criteria of having both a
frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a
GC Content >0.5.

20

EXAMPLE 1

CpG Island Hypermethylation Increased with the Progression of EAC

This Example shows the results of an analysis of the methylation status of a panel of
CpG islands associated with 19 different genes selected for their known involvement in
25   carcinogenesis or because they have been shown to be methylated in other tumors (*see* Table
1, and under "Definitions," above), and of one non-CpG island sequence (*MTHFR* control
sequence), for a total of 20 gene loci.

Quantitative methylation data of the 20 genes from a screen of 84 tissue specimens
from 31 patients with different stages of Barrett's esophagus and/or associated
30   adenocarcinoma showed a general increase in the frequency and in the quantitative level of
CpG island hypermethylation at progressively advanced stages of disease. Accordingly,
genes were grouped into distinct classes by their methylation behavior, based on both
frequency and level of hypermethylation in various tissues (Figure 1).

35   Materials and Methods

Sample Collection and histopathologic examination. Multiple tissue samples (normal

27

esophagus (NE), normal stomach (S), intestinal metaplasia (IM), dysplasia (DYS) and/or adenocarcinoma (T)) from a total of 51 patients (range 39-86 years of age) with either adenocarcinoma or IM as the most advanced stage of disease were collected.

5   The initial set of samples analyzed included biopsies from 31 patients which were collected fresh and subdivided such that a part of each specimen was immediately frozen in liquid nitrogen and also embedded in paraffin for histopathologic examination by a pathologist (K.W.). Normal esophageal tissue was collected from every patient 10 cm or more away from the diseased areas. Frozen section examination of the frozen tissues was performed if the diagnosis was uncertain. The site of origin of the cancers was classified as

10   esophageal if the epicenter of the tumor was above the anatomic gastroesophageal junction, with the junction defined as the proximal margin of the gastric rugal folds. TNM staging was used to classify the stage of each adenocarcinoma.

A second set of samples were obtained for a follow-up study of 20 cases. Two groups of IM samples were collected: patients that had only IM as the most advanced stage of disease

15   (8 patients), and patients that had IM with associated dysplasia/adenocarcinoma located in another region of the esophagus (12 patients). H&E slides (5-micron sections) for each sample were prepared and examined by a pathologist (K.W.) to verify and localize the IM tissue. Cases that showed any signs of dysplasia or adenocarcinoma in the paraffin block used for analysis were excluded from this follow-up study. The IM tissues were carefully

20   microdissected away from other cell types from a 30-micron section adjacent to the 5-micron H&E section. All specimens were classified according to the highest grade histopathologic lesion present in that sample. Approval for this study was obtained from the Institutional Review Board of the University of Southern California Keck School of Medicine.

Nucleic Acid Isolation. Genomic DNA was isolated from the frozen tissue biopsies by

25   a simplified proteinase K digestion method (Laird et al., *Nucleic Acids Res.* 19:4293, 1991). The DNA from the paraffin tissues was extracted in lysis buffer (100 mM Tris-HCl, pH 8; 10 mM EDTA; and 1mg/ml Proteinase K) overnight at 50°C (Shibata et al., *Am. J. Pathol.* 141:539-543, 1992).

Sodium Bisulfite Conversion. Sodium bisulfite conversion of genomic DNA was

30   performed as previously described (Olek et al., *Nucleic Acids Res.* 24:5064-5066, 1996). The beads were incubated for 14 hours at 50°C to ensure complete conversion. Sodium bisulfite treatment converts unmethylated cytosines to uracil, while leaving methylated cytosine residues intact (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-31, 1992).

MethyLight™ Analysis. After sodium bisulfite conversion, the methylation analysis

35   was performed by the fluorescence-based, real-time PCR assay MethyLight™, as described herein, and as previously described (Eads et al., *Cancer Res.* 60:5021-5026, 2000; Eads et al.,

28

Cancer Res. 59:2302-2306, 1999; Eads et al., *Nucleic Acids Res.* 28:E32, 2000). Two sets of primers and probes, designed specifically for bisulfite converted DNA, were used: a methylated set for the gene of interest and a reference set, beta-actin (*ACTB*) to normalize for input DNA. Specificity of the reactions for methylated DNA were confirmed separately using

5   human sperm DNA (with very low levels of CpG island methylation) and *SssI* (New England Biolabs)-treated sperm DNA (heavily methylated) as previously described (Eads et al., *Cancer Res.* 60:5021-5026, 2000).

The percentage of fully methylated molecules at a specific locus was calculated by dividing the *GENE/ACTB* ratio of a sample by the *GENE/ACTB* ratio of *SssI*-treated sperm

10   DNA and multiplying by 100. The abbreviation PMR (Percent of Methylated Reference) is used to indicate this measurement. The methylation analysis on the paraffin microdissected samples was performed following bisulfite treatment as described above by an investigator blind to the associated dysplasia status of the samples.

TABLE II lists the MethyLight™ primer and probe sequences (SEQ ID NOs:1-65),

15   based on Genbank sequence data (except for SEQ ID NOs:64 and 65, see below), used in the present methylation analysis. Three oligos were used in every reaction: two locus-specific PCR primers flanking an oligonucleotide probe with a 5' fluorescent reporter dye (6FAM) and a 3' quencher dye (TAMRA) (Livak et al., *PCR Methods Appl.* 4:357-362, 1995). The Genbank accession number for each sequence is listed with the corresponding PCR amplicon

20   location within that sequence. The %GC content, CpG observed/expected value and CpG:GpC ratio of 200 base pairs encompassing the MethyLight amplicon are indicated for each gene. The reaction type is designated "M" for methylation reaction and "C" for control reaction. The bisulfite treated DNA strand (top ("T") or bottom ("B")) and amplicon orientation (parallel ("P") or antiparallel ("A")) is also indicated. All primer and probe

25   sequences are listed in the 5' to 3' direction. The numbers in brackets after each primer or probe sequence correspond to the associated SEQ ID NOs. The single asterisk (*) notes that there are two bases in our *CDKN2A* primers that differ from this GenBank sequence, since a preliminary high-throughput GenBank entry was the only available sequence at the time of applicants' primer design. The correct primers should be the following: forward,

30   TGGA<u>G</u>TTTTCGGTTGATTGGTT (SEQ ID NO:64) and reverse, AACAACG<u>C</u>CCGCACCTCCT (SEQ ID NO:65). The bases differing from the GenBank sequences are underlined. The double asterisk (**) indicates that the start site is not well defined.

TABLE II. MethyLight™ primer and probe sequences

HUGO Gene Name	Genbank Accession Number	Bisulfite (T or B Strand)	Amplicon Location (Genbank Numbering)	Amplicon Location (bp) Relative to Transcription Start	% GC Content	Obs/Exp CpG	CpG:gpC	Forward Primer Sequence (5'-3')	Probe Sequence (5'-3')	Reverse Primer Sequence (5'-3')
"Class A" genes										
CDKN2A*	NM_000077	T	66-133bp	+9/+69	72	1.68	0.81	TGGAATTTCG GTTGATTGGTT [1]	6FAM-ACCCGACCCGAACCGC G-TAMRA [3]	AACAACGTCCG CACCTCCT [2]
ESR1	X62462	T	2784bp-2884bp	+14/+114	69	1.52	0.83	GGCGTTCGTTT TGGGATTG [4]	6FAM-CGATAAAACCGAACGAC CCGACGA-TAMRA [6]	GCCGACACGCG AACTCTAA [5]
MYOD1	AF027148	T	9889bp-9962bp	-375/-302	61	1.12	0.74	GAGCGCGCGT AGTTAGCG [7]	6FAM-CTCCAACACCCGA CTACTATATCCGCGAAA-TAMRA [9]	TCCGACACGCC CTTTCC [8]
"Class B" genes										
CALCA	X15943	T	1706bp-1806bp	-128/-28	60	0.72	0.53	GTTTTGGAAGT ATGAGGGTGAC G [10]	6FAM-ATTCCGCCAATAC ACAACAACCAATAAACG-TAMRA [12]	TTCCCGCCGCTA TAAATCG [11]
MGMT	U95038	B	206bp-297bp	+392/+483	67	0.88	0.55	CTAACGTATAA CGAAAATCGTA ACAACC [16]	6FAM-CCTTACCTCTAAAT ACCAACCCCAAACCCG-TAMRA [18]	AGTATGAAGGG TAGGAAGAATT CGG [17]
TIMP3	U33110	T	1051bp-1143	+1051/+1143	78	1.04	0.63	GCGGTCGGAGGT TAAGGTTGTT [13]	6FAM-AACTCGCTCGCCCGCCGA A-TAMRA [15]	CTCTCCAAAATT ACCGTACGCG [14]
"Class C" genes										
APC	U02509	B	759bp-8322bp	+7/+80	69	1.36	0.71	GAACCAAAACG CTCCCCAT [19]	6FAM-CCGGTCGAAAACCCGCC GATTA-TAMRA [21]	TTATATGTCGGT TACGTGCGTTTA TAT [20]

HUGO Gene Name	Genbank Accession Number	Bisulfite (T or B) Strand)	Amplicon Location (Genbank Numbering)	Amplicon Location (bp) Relative to: Transcription Start	% GC Content	Obs/Exp CpG	CpG:gpC	Forward Primer Sequence (5'-3')	Probe Sequence (5'-3')	Reverse Primer Sequence (5'-3')
"Class D" genes										
ARF	AF082338	T	5447bp-5515bp	-203/-135	78	2	0.89	ACGGGGCGTTTCGGTAGTT [22]	6FAM-CGACTCTAAACCCTACGCACGGCGAAA-TAMRA [24]	CCGAACCTCCAAAATCTCGA [23]
CDH1	L34545	T	842bp-911bp	-180/-111	69	1.2	0.68	AATTTTAGGTTAGAGGGGTTATCGCGT [25]	6FAM-CGCCCACCCGACCTCGCAT-TAMRA [27]	TCCCCAAAACGAAACTAACGAC [26]
CDKN2B	S75756	T	350bp-430bp	+222/+302	64	1.44	0.82	AGGAAGGAGAGAGTGCGTCG [28]	6FAM-TTAACGACACTCTTCCCTTCTTCCCACG-TAMRA [30]	CGAATAATCCACCGTTAACCG [29]
GSTP1	M24485	T	1146bp-1245bp	-79/+21	75	2	0.93	GTCGGCGGTCGTGATTTAGTATTG [31]	6FAM-AAACCTCGGCGACCTCCGAACCTTATAAAA-TAMRA [33]	AAACTACGACGACGAAACTCCAA [32]
MLH1	U26559	T	254bp-341bp	-662/-575	68	2.08	1.04	CGTTATATATCGTTCGTAGTATTCGTGTTT [35]	6FAM-CGGCGACGTCAAACGCCACTACG-TAMRA [36]	CTATCGCCGCCTCATCGT [34]
PTGS2	AF044206	T	6779bp-6924bp	-362/-217	62	0.8	0.71	CGGAAGCGTTCGGGTAAAG [37]	6FAM-TTTCCGCCAAATATCTTTCTTCTTGCA-TAMRA [39]	AATTCCACCGCCCAAAC [38]
THBS1	J04835	B	1642bp-1716bp	-636/-562	75	2.08	1	CGACGCACCAACCTACCG [40]	6FAM-ACGCCCGGCGCTCACCTCCC-T-TAMRA [42]	GTTTGAGTTGGTTTTACGTTCGTT [41]
"Class E" genes										
CTNNB1	X89448	T	583bp-664bp	-413/-332	75	1.92	1	GGAAAGGCGCGTCGAGT [43]	6FAM-CGGCGGTTCCCGAACCG-TAMRA [45]	TCCCCTATCCAAACCCG [44]

HUGO Gene Name	Genbank Accession Number	Bisulfite (T or B) (Strand)	Amplicon Location (Genbank Numbering)	Amplicon Location (bp) Relative to Transcription Start	% GC Content	Obs/Exp CpG	CpG:GpC	Forward Primer Sequence (5'-3')	Probe Sequence (5'-3')	Reverse Primer Sequence (5'-3')
RB1	L11910	T	1781bp-1900bp	-279/-160	68	2.24	1.27	TTAGTTCGGCGTATCGATTAGCG [46]	6FAM-TCACGTCCGGCGAAACTCCCGA-TAMRA [48]	ACTAAAACGCCGCGTCCAA [47]
TGFBR2	U52240	T	256bp-333bp	-58/+19	75	2.08	0.81	GCGCGGGAGCGTAGTTAGG [49]	6FAM-CACGAACGACGCCTTCCCGAA-TAMRA [51]	CAAACCCCGCTACTCGTCAT [50]
TYMS	D00517	T	10052bp-1128bp	+301/+377	75	2.16	0.9	CGGCGTTAGGAAGGACGAT [52]	6FAM-CCGAATACCGACAAAATACCGATACCCGT-TAMRA [54]	TCTCAAACTATAACGGCGCCTACAT [53]
"Class F" genes										
HIC-1	L41919	T	562bp-662bp	-55/+45**	71	1.12	0.5	GTTAGGGCGGTTAGGGCGGTC [58]	6FAM-CAACATCGTCTACCCAACACACTCTCCTACG-TAMRA [60]	CCGAACGCCTCCATCGTAT [59]
"Class G" genes										
MTHFR	AF10597 7	T	1123bp-212bp	+72/+45	57	0.56	0.54	TGGTAGTGAGAGTTTTAAAGATAGTTCGA [55]	6FAM-TCTCATACCGCTCAAAATCCAAACCCG-TAMRA [57]	CGCCTCATCTTCTCCCGA [56]
"Control" gene										
ACTB	Y00474	T	390bp-522bp	-1,599/-1,467**	58	0.16	0.14	TGGTGATGGAGGAGGTTTAGTAAGT [61]	6FAM-ACCACCACCCAACACACAATAACAAACACA-TAMRA [63]	AACCAATAAAACCTACTCCTCCCTTAA [62]

TABLE II cont. MethyLight™ primer and probe sequences. Numbers in brackets correspond to SEQ ID NOs: 1-63.

‡ Gene "Classes" are defined according to the present invention.

Statistics. The PMR values obtained by MethyLight™ (see above) were "dichotomized" at 4 PMR for statistical purposes as described previously (Eads et al., *Cancer Res*. 60:5021-5026, 2000. Dichotomization facilitates graphical representation, and moderates the quantitative impact of gene loci with different levels of hypermethylation,

5   resulting in a more reliable cross-gene comparison of hypermethylation frequencies. Specifically, dichotomization equalizes the quantitative impact of methylated genes within each class (*see* "Epigenetic gene classes," below), simplifying cross-gene comparisons of methylation frequencies.

       A dichotomization point of 4 PMR was selected because it gave the best

10   discrimination between normal and malignant tissues, across the board for all CpG islands (Eads et al., *Cancer Res*. 60:5021-5026, 2000). However, the precise dichotomization point does not significantly affect the statistics or alter the conclusions, and other dichotomization points are within the scope of the present invention (see below).

       Accordingly, samples containing 4 PMR or higher were designated as methylated and

15   given a value of 1, while samples containing less than 4 PMR were designated as unmethylated and given a value of 0. The cumulative value of genes methylated in each class (*see* Epigentic gene classes" A-G, herein below), or for all 19 genes was then used as a continuous variable in a Fisher's Protected Least Significant Difference test, adapted for use with unequal sample sizes (SAS Statview software) to obtain *p*-values. The different

20   parameters such as tissue type, presence of associated dysplasia, tumor stage, etc., were used as the nominal variables. The IM samples in the above-mentioned "follow-up" study of hypermethylation in IM, and the presence of associated dysplasia and/or carcinoma, were further dichotomized at 1 or fewer, versus two or more Class A genes methylated. A Fisher's exact test was then used to determine statistical significance.

25

       Results

       CpG Island Hypermethylation and the Progression of EAC. The methylation status of a panel of CpG islands associated with 19 different genes and of one non-CpG island sequence for a total of 20 gene loci, was analyzed by the quantitative, high-throughput

30   MethyLight™ assay (Eads et al., *Cancer Res*. 59:2302-2306, 1999; Eads et al., *Nucleic Acids Res*. 28:E32, 2000). The efficiencies of the methylation reactions were controlled for in each analysis by including unmethylated control DNA and methylated control DNA (Eads et al., *Cancer Res*. 60:5021-5026, 2000). The 20 genes were selected for their known involvement in carcinogenesis or because they have been shown to be methylated in other tumors (*see*

35   Table 1, and under "Definitions," above). We included a region located in the *MTHFR* gene as a "non-CpG island" control for a single copy sequence that does not satisfy the criteria (*see*

"Definitions," above) of a CpG island. CpG dinucleotides outside of an island are presumably normally methylated, unlike CpG dinucleotides within CpG islands.

Figure 1 illustrates the quantitative methylation data of the 20 genes from our screen of 84 tissue specimens from 31 patients with different stages of Barrett's esophagus and/or

5   associated adenocarcinoma. Methylation analysis was performed using the MethyLight assay (Eads et al., *Cancer Res.* 59:2302-2306, 1999; Eads et al., *Nucleic Acids Res.* 28:E32, 2000). The percentage of fully methylated molecules at a specific locus (PMR = Percent of Methylated Reference) was calculated by dividing the *GENE/ACTB* ratio of a sample by the *GENE/ACTB* ratio of *SssI*-treated sperm DNA and multipling by 100. The resulting

10   percentages were then dichotomized at 4% PMR to facilitate graphical representation and to reveal tissue-specific patterns. The various squares, each having one of four possible shading intensity levels (see bottom axis of Figure 1), designate samples with less than 4 PMR, 4 - 20 PMR, 21 - 50 PMR and more than 51 PMR, where progressively increasing shading intensity levels correspond to progressively higher PMR values. The tissue types are shown on the left.

15   The TNM tumor staging is designated by "1", "2", "3" and "4". The occurrence of distally located dysplasia and/or adenocarcinoma in the patient is indicated at the right of the figure by "YES" if present and "NO" if absent. "N" indicates an analysis for which the control gene *ACTB* did not reach sufficient levels to allow the detection of a minimal value of 1 PMR for that methylation reaction in that particular sample.

20   There was a general increase in the frequency and in the quantitative level of CpG island hypermethylation at progressively advanced stages of disease. However, the propensity for aberrant methylation of the genes was not uniform. Genes differed both in their frequency and in their levels of hypermethylation in various tissues.

Therefore, according to the present invention, genes can be grouped into classes based

25   on their methylation behavior (Classes A-G, as shown at the right of Figure 1). This allowed for a visual assessment of concordant methylation of the different genes during various stages of turmorigenesis. A rationale for each of the gene classes is presented in the following section.

30   *Epigenetic Gene Classes*. The analysis of combined behavior of genes with different levels of DNA methylation would, without appropriate data treatment, be expected to lead to a *bias* of the group behavior towards genes with quantitatively high levels of DNA methylation. For instance, the mean values for gene "Class B" for most of the tumor samples would be driven primarily by the *TIMP3* values, since this gene tended to have higher levels

35   of methylation than the other two genes in this group (*see* Figure 1).

Therefore, the methylation values used to generate Figure 1 were collapsed into a

34

binary variable with a dichotomization point of 4 PMR to equalize the quantitative impact of methylated genes within each epigenetic class. Samples containing 4 PMR or higher were designated as methylated and given a value of 1, while samples containing less than 4 PMR were designated as unmethylated and given a value of 0 (*see* "*Statistics*" above, under

5   "*Materials and Methods*"). This dichotomization moderates the effect of highly methylated genes, simplifies cross-gene comparisons of methylation frequencies, as shown in Figure 2, and allowed the calculation of class averages of methylation frequencies as shown in Figure 3 (below).

        Figure 2 shows the percent of samples methylated for each gene by tissue type. The

10   data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. The genes, according to the present invention, were grouped according to their respective epigenetic gene classes (A-G) as shown in Figure 1. The letter "n" equals the number of samples analyzed for each tissue.

        The suitability of the 4 PMR dichotomization point was based on its ability to

15   discriminate between the different tissue types, as shown in Figures 1-3 (*see also* Klump et al., *Gastroenterology*. 115:1381-1386, 1998). Other dichotomization point values are within the scope of the present invention, where such dichotomization point values moderate the statistical effects of highly methylated genes, simplify cross-gene comparisons of methylation frequencies, and facilitate calculation of class averages of methylation frequencies. For

20   instance, there is still a statistically significant difference in the mean percent of genes methylated (out of 19 genes) between the normal esophageal mucosa and the IM ($p = 0.0003$), DYS ($p < 0.0001$) and T ($p < 0.0001$) tissues when the data is dichotomized at 10 PMR.

        Additionally, all of the statistically significant findings of the NE and IM methylation frequency with or without associated dysplasia (*see* Example 3, below) remain significant at a

25   dichotomization point of 10 PMR, instead of 4 PMR. It is important to note that 4 PMR is not comparable to a 4% methylation level of a single CpG dinucleotide. Rather, it indicates that in this sample, 4% of the DNA molecules had *complete* methylation at all CpG dinucleotides covered by the three MethyLight™ primers (usually about 8 CpGs). The nature of the MethyLight™ assay is such that it is oblivious to all other methylation patterns that may be

30   present (Eads et al., *Nucleic Acids Res*. 28:E32, 2000).

        Therefore, 4 PMR is likely to represent a higher mean level of methylation than 4%. The extensively methylated molecules that are assayed by MethyLight™ are likely to represent alleles that have been completely silenced by CpG island hypermethylation, although this was not investigated herein.

35   Of the panel of 20 genes, the most informative genes were those with an intermediate frequency of hypermethylation (ranging from 15% (*CDKN2A*) to 60% (*MGMT*) of the sample

values above the 4 PMR methylation cutoff). This group was further subdivided into three epigenetic gene classes according to the absence (Class "A") or presence (Class "B") of methylation in normal esophageal mucosa and stomach, or the infrequent methylation of normal esophageal mucosa accompanied by methylation in all normal stomach samples (Class

5   "C"). The other genes were less informative, since the incidence of hypermethylation was either very infrequent (Class "D"), completely absent (Class "E"), or ubiquitous (Classes "F" and "G") regardless of tissue type (Figures 1, 2 and 3).

Epigenetic gene Class A comprises the genes *CDKN2A*, *ESR1* and *MYOD1* (Figures 1, 2 and 3). There was a statistically significant difference in the methylation frequency of *ESR1*

10   ($p = 0.0001$) and *MYOD1* ($p = 0.0038$) of normal esophagus (NE), as compared to IM tissue, but not for *CDKN2A* ($p = 0.097$). The frequency of *CDKN2A* methylation increased significantly in the more advanced stages of the adenocarcinoma (T) ($p < 0.0001$).

Epigenetic gene Class B comprises the genes *CALCA*, *MGMT* and *TIMP3*. In contrast to Class A, this class exhibited methylation in the normal esophageal mucosa (NE) and

15   stomach (S) tissue (Figures 1 and 2). Only *TIMP3* showed a significant difference in methylation frequency between the NE and IM values ($p = 0.0074$).

Epigenetic gene Class C comprises the gene *APC* which was, in contrast to genes of Classes A and B, methylated in all normal stomach samples (Figures 1 and 2). This confirms previous documentation of *APC* methylation in normal stomach tissue (Eads et al., *Cancer*

20   *Res.* 60:5021-5026, 2000). The mechanism which protects *APC* from methylation in the normal esophageal tissues (NE) but not in normal stomach tissues (S) is not clear.

Epigenetic gene Class D comprises the genes *ARF*, *CDH1*, *CDKN2B*, *GSTP1*, *MLH1*, *PTGS2* and *THBS1*, which were infrequently methylated (Figures 1 and 2). There was a slight increase in the frequency of this class of genes in adenocarcinoma (T), but this did not

25   approach statistical significance (Figure 3). Interestingly, with the exception of *PTGS2*, which has not yet been investigated in other systems, the remaining Class D genes are frequently hypermethylated in other tumor types (Table 2).

Epigenetic gene Class E comprises the *CTNNB1*, *RB1*, *TGFBR2* and *TYMS1* genes, which were unmethylated at each stage in the progression of EAC. Similar to most Class D

30   genes, *RB1* and *TGFBR2* have been found to be hypermethylated in other tumors types (*see* Table 1, and literature references under "DEFINITIONS" herein above). It should be noted that all samples scored postitive for DNA input as measured by the control gene (*ACTB*). Therefore, the lack of detectable DNA methylation cannot be attributed to a lack of input DNA. The control reaction was sufficient in each sample, so that a level as low as 1 PMR for

35   a given test gene could be detected. The integrity and specificity of all methylation reactions was confirmed using *in vitro* methylated human DNA.

36

The epigenetic Class F comprises the *HIC1* gene, which was completely methylated, regardless of tissue type (Figures 1 and 2). *HIC1* is commonly methylated in other types of cancers (Jones & Laird, *Nat. Genet.* 21:163-167, 1999; Baylin & Herman, *Trends Genet.* 16:168-174, 2000), and has been shown to be methylated in normal breast ductal tissue and

5   bone marrow samples of breast cancer and AML patients, respectively (Melki et al., *Cancer Res.* 59:3730-3740, 1999; Fujii et al., *Oncogene.* 16:2159-2164, 1998). Nevertheless, the finding of ubiquitous methylation of a CpG island in normal tissues was unexpected. Therefore, the validity of the *HIC1* MethyLight™ results was confirmed using a different technique (HpaII-PCR) (Singer-Sam et al., *Nucleic Acids Res.* 18:687, 1990).

10   Epigenetic Class G comprises the non-CpG island *MTHFR* gene, used herein as a control. Interestingly, the ubiquitous *HIC1* methylation pattern is similar to the non-CpG island *MTHFR* control (Class G), however the percentage of methylated molecules was quantitatively higher for *HIC1* (Figure 1).

15   *Epigenetic Profiles of EAC Progression.* Each tissue type showed a unique epigenetic profile or fingerprint that changed during disease progression (Figure 3, upper panel).

Figure 3 shows a comparison of epigenetic profiles according to the present invention. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. Error bars represent the standard error of the mean. *Upper*

20   *panel*: Mean percent of genes methylated in each gene Class (A-F or ALL 19 CpG islands) by tissue type (N, normal esophagus; S, stomach; IM, intestinal metaplasia; DYS, dysplasia; T, adenocarcinoma). The error bars represent the standard error of the mean (SEM). *Lower panel*: Statistical analysis of the difference in mean percent of genes methylated in different tissues by gene Class (A-F) or for all 19 CpG islands combined (ALL). The *p*-values were

25   generated by a Fisher's Protected Least Significant Difference (PLSD) test, adapted for use with unequal sample numbers (SAS Statview™ software).

Classes A, B and C were methylated at a significantly higher frequency in IM tissue than in normal esophageal mucosa (NE) (Figure 3, upper and lower panels). Furthermore, the transition from IM to dysplasia (DYS) or malignancy (T) was associated with an additional

30   increase in Class A methylation (Figure 3, upper and lower panels). The lack of a significant difference between dysplasia and adenocarcinoma for any of the gene classes or when all 19 genes are combined (Figure 3, upper and lower panels) suggests that most of these abnormal epigenetic alterations occur early in the progression of EAC.

35   *In summary of this Example.* According to the present invention, quantitative methylation data of 20 genes (Tables I and II, above) from a screen of 84 tissue specimens

from 31 patients with different stages of Barrett's esophagus and/or associated adenocarcinoma showed a general increase in the frequency and in the quantitative level of CpG island hypermethylation at progressively advanced stages of disease (Figures 1-3, above).

5   Additionally, genes were grouped into novel epigenetic classes based on their methylation behavior (Classes A-G, as shown herein in Figures 1-3) during tumor progression. This allowed for graphical representation of concordant methylation of the different genes during various stages of turmorigenesis, which can be readily appreciated by means of a simple visual assessment.

10   Each tissue type showed a unique epigenetic profile or fingerprint that changed during disease progression (Figure 3, upper panel). Classes A, B and C were methylated at a significantly higher frequency in IM tissue than in normal esophageal mucosa (NE) (Figure 3, upper and lower panels). Furthermore, the transition from IM to dysplasia (DYS) or malignancy (T) was associated with an additional increase in Class A methylation (Figure 3,

15   upper and lower panels).

EXAMPLE 2

Hypermethylation was Reflective of EAC Tumor Grade and Stage

           This Example examines whether the grade or stage of an esophageal adenocarcinoma
20   correlates with a higher frequency of CpG island hypermethylation. According to the present invention, for EAC, epigenetic Class A gene methylation is significantly higher in stage II, III and IV tumors relative to less advanced stage I tumors (Figure 4).

Materials and Methods

25   *TNM staging.* The American Joint Committee on Cancer ("AJCC") has designated staging by TNM classification (Tumor; lymph Node metastasis, distant Metastasis). TNM staging was used to classify the stage of each esophageal adenocarcinoma from the tissues of Example 1.

           Methylation and statistical analysis. Methylation and statistical analysis was as
30   described herein under Example 1.

Results

           Methylation of epigenetic Class A genes increases with tumor stage. Moderately differentiated tumors have significantly less frequent Class A methylation compared to poorly
35   differentiated tumors ($p = 0.045$). Additionally, Figure 4 (upper and lower panels) shows that there is a significantly higher mean number of Class A genes methylated in stage II, III and

IV tumors relative to less advanced stage I tumors. The differences between stage I tumors and stage II, III and IV tumors did not reach statistical significance for any of the other epigenetic gene classes.

Figure 4 shows the relationship between Class A methylation frequency and tumor
5   stage according to the present invention. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. *Upper panel*: Mean number of genes methylated for Class A with respect to tumor stage (I-IV) is shown (see Figure 1). The error bars represent the standard error of the mean (SEM). The letter "n" equals the number of samples analyzed in each tumor stage. *Lower panel*: Statistical analysis
10 of the difference in mean number of Class A genes methylated by tumor stage. The *p*-values were generated by a Fisher's Protected Least Significant Difference (PLSD) test, adapted for use with unequal sample numbers (SAS Statview[TM] software).

In summary for this Example. According to the present invention, in addition to the
15 epigenetic profiles or fingerprints (comprising the gene classes disclosed herein) that can be used to assess oncogenic progression, the mean number of methylated Class A genes can be used to assess the relative stages of EAC tumors.

EXAMPLE 3

20 Methylation of Premalignant Tissues With or Without Associated Dysplasia

This Example shows that the frequency of Class B methylation in the normal esophagus (NE) was found to be significantly higher in patients with associated dysplasia/tumor ($p = 0.0037$) (Figure 1). Additionally, Class A methylation was found to be more frequent in IM samples from patients with concurrent dysplasia or cancer, than in IM
25 samples from patients without any evidence of further progression ($p < 0.0001$) (Figures 1 and 5). That is, there was a significant positive association between hypermethylation of epigenetic Class A genes in IM tissue, and the presence of associated dysplasia or cancer (Figure 5).

30 Materials and Methods

Histopathology. Histopathological classification was as described under "Materials and Methods," Example I above.

Methylation and statistical analysis. Methylation and statistical analysis was as described herein under Example 1.

35

Results

39

Methylation of Premalignant Tissues with or without Associated Dysplasia. The occurrence, according to the present ivention, of CpG island hypermethylation in some cases of IM for Class A and some cases of normal esophageal mucosa for Class B raised the question whether these methylation events represent normal methylation patterns in these

5   non-dysplastic tissues, or whether they reflect methylation changes that predispose cells to further progression. In the latter case, one would expect to find a higher frequency of such CpG island hypermethylation in these tissues in patients who have already undergone further disease progression. Therefore, the frequency of such CpG island hypermethylation was compared between tissues (of the present study) with or without associated dysplasia.

10   In the initial study, patients were divided based on whether or not they had Barrett's esophagus (IM) as their most advanced stage of disease (Figure 1, "NO") or whether they had associated dysplasia and/or adenocarcinoma present in a different region of the esophagus (Figure 1, "YES"). The frequency of Class B methylation in the normal esophagus (NE) was indeed found to be significantly higher in patients with associated dysplasia/tumor ($p =$

15   0.0037) (Figure 1). Additionally, Class A methylation was found to be more frequent in IM samples from patients with concurrent dysplasia or cancer, than in IM samples from patients without any evidence of further progression ($p < 0.0001$) (Figure 1).

A potential criticism of this analysis is that the same set of samples was used to delineate the class of genes, as was used to test the association with a clinical parameter.

20   Therefore, a follow-up study of 20 additional cases of IM was performed entirely independent of the first data set.

In the follow-up study of 20 cases, two groups of IM samples were collected: patients that had only IM as the most advanced stage of disease (8 patients), and patients that had IM with associated dysplasia/adenocarcinoma located in another region of the esophagus (12

25   patients). H&E slides (5-micron sections) for each sample were prepared and examined by a pathologist (K.W.) to verify and localize the IM tissue. Cases that showed any signs of dysplasia or adenocarcinoma in the paraffin block used for analysis were excluded from this follow-up study. The IM tissues were carefully microdissected away from other cell types from a 30-micron section adjacent to the 5-micron H&E section. All specimens were

30   classified according to the highest grade histopathologic lesion present in that sample.

The initial study had revealed that all IM samples associated with further disease progression ("YES") had at least two Class A genes methylated, while all IM samples without associated dysplasia or adenocarcinoma ("NO") did not show any methylation of Class A genes (Figure 1, under "Barrett's (IM)"). Therefore, a state of having two or more Class A

35   genes methylated was defined as an indicator of increased risk for the presence of associated dysplasia or cancer.

<div align="center">40</div>

The data from our first series gave a p-value of 0.0048 in a Fisher's exact test of this association (Figure 5, left panel). The follow-up series of 20 independent cases gave a p-value of 0.018 (Figure 5, right panel).

Figure 5 shows the percent of two or more Class A genes methylated in intestinal

5   metaplasia ("IM") tissues with ("Y"), or without ("N") associated dysplasia and/or adenocarcinoma. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. *Left panel*: Class A methylation in the IM data illustrated in Figure 1. Right panel: Class A methylation in the IM for a completely independent follow-up study of twenty different microdissected IM samples. The error bars

10   represent the standard error of the mean (SEM). The letter "n" equals the number of samples analyzed in each tissue group.

Therefore, the positive association between hypermethylation of Class A genes and the presence of associated dysplasia or cancer is significant. It should be noted that the IM samples without associated dysplasia in this follow-up study (Figure 5, right panel) showed a

15   *low* frequency of samples with at least two genes methylated, which is in contrast to the *absence* of methylation in the first study (Figure 1, and Figure 5, left panel). This may be attributed to the fact that the samples in the second series were microdissected from paraffin sections. Therefore, there is a lower background of unmethylated stromal cells in the sample. In this case, the methylation signal is not as diluted by other normal cells and consequently

20   the ratio of methylated molecules to total DNA may rise above the 4 PMR threshold. Alternatively, dysplastic or malignant tissue may have been missed during the endoscopic survey in some of the cases scored as free of further disease progression due to the sampling limitations of endoscopy. This is a well-documented problem in the detection of esophageal adenocarcinoma (Peters et al., *J. Thorac. Cardiovasc. Surg.* 108:813-821, 1994).

25

EXAMPLE 4

No Clear Evidence of CpG Island Methylator Phenotype ("CIMP") for EAC

This Example shows that, for the present study of EAC, there was no clear evidence of a separate group of CIMP tumors, as has been previously defined for colorectal and gastric

30   cancer (Toyota et al., *Proc. Natl. Acad. Sci. U S A.* 96:8681-8686, 1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999). However, CpG island hypermethylation in EAC did occur across multiple loci in a given sample. Furthermore, the number of loci hypermethylated in a single sample increased as the disease progressed through different histological stages (Figure 6). The bimodal distributions seen in IM tissues (Figure 6) can be fully attributed to the

35   concurrent association with dysplasia or cancer described herein above.

Materials and Methods

      Histopathology. Histopathological classification was as described under "Materials and Methods," Example I above.

      Methylation and statistical analysis. Methylation and statistical analysis was as

5   described herein under Example 1.

Results

      CIMP Analysis. It has previously been reported that a subset of colorectal and gastric tumors display a CpG island methylator phenotype ("CIMP"), characterized by widespread,

10   aberrant hypermethylation changes affecting multiple loci in a single tumor (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686, 1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999). This is reflected in a *bimodal* distribution of the frequency of the number of genes methylated in a group of tumors (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686, 1999). CIMP tumors are a distinct group of tumors that are defined by a high degree of

15   concordant CpG island hypermethylation of genes exclusively methylated in cancer, or "type-C" genes. CIMP is currently thought to be a new, distinct, yet major pathway of tumorigenesis (Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686, 1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999).

      Therefore the question of whether esophageal adenocarcinoma tumors exhibit a CpG

20   island methylator phenotype (CIMP) was investigated.

      Class A genes of the present invention most closely exemplify the "type-C" genes, because they lack methylation in the normal tissues. The distribution of the number of Class A genes methylated was examined for EAC (Figure 6).

      Figure 6 shows, according to the present invention, methylation frequency

25   distributions in the progression of esophageal adenocarcinoma. The data was dichotomized at 4 PMR, with 4 PMR and higher designated as methylated, and below 4 PMR as unmethylated. The proportion of patients with zero to three (Class A), zero to nine (Classes A + D) and zero to fourteen CpG islands (Classes A + B +C + D) methylated in each tissue is shown. Class E and F CpG islands were not included since there was no variation in the

30   frequency of methylation between the different tissue. The letter "n" equals the number of samples analyzed in each tissue.

      However, the frequency of genes methylated in the adenocarcinoma tissue did not show the expected bimodal distribution of CIMP (Figure 6) (Toyota et al., *Cancer Res.* 59:5438-5442, 1999). Similar results were observed when Class D genes, which also exhibit

35   type C methylation, were included along with Class A (Figure 6, middle panel) and when Classes A, B, C and D genes were combined (Figure 6, right panel). Classes E and F genes

were not included since they did not exhibit any methylation variation between the different tissue types.

There was a single sample with 10 out of 14 Class A-D genes methylated (Figure 1, Case #3 and Figure 6). However, this sample only stands out when Class B genes, which are

5   methylated in normal esophageal mucosa and therefore do not satisfy the definition of "type-C" genes that constitute the CIMP phenotype, are included.

Therefore, there was no clear evidence of a separate group of CIMP tumors in the present study of esophageal adenocarcinoma, as has been previously defined for colorectal and gastric cancer.

10   However CpG island hypermethylation in EAC did occur across multiple loci in a given sample. Furthermore, the number of loci hypermethylated in a single sample increased as the disease progressed through different histological stages (Figure 6). The bimodal distributions seen in IM tissues (Figure 6) can be fully attributed to the concurrent association with dysplasia or cancer described herein above.

15

EXAMPLE 5

Array- and Microarray-based Applications

Microarray-based embodiments are within the scope of the present invention. For example, one such array-based embodiment uses differential methylation hybridization

20   ("DMH"), (Huang et al., *Hum. Mol. Genet.*, 8:459-470, 1999; Yan et al., *Clin. Cancer Res.* 6:1432-38, 2000). DMH is applied to screen paired test and normal samples and to determine whether patterns (*see* "Epigenetic patterns," herein under Example 1) of specific epigenetic alterations correlate with pathological parameters in the tissue samples analyzed. "Amplicons" (*Id*), representing a pool of methylated CpG DNA derived from these samples,

25   are used as hybridization probes in an array panel containing the CpG island tags of the present invention.

Accordingly, one or more of the CpG island sequences associated with 19 of the 20 disclosed gene sequences (*i.e., APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2*

30   and *TYMS* (*see* TABLES I and II, above; *and see* under "Definitions," above), or methylation-altered DNA sequence embodiments thereof, can be used as CpG island tags in an array or microarray-based assay embodiment. These 19 gene sequence regions are defined herein by the oligomeric primers and probes corresponding to SEQ ID NOs:1-54, 58-60, 64 and 65 (*see* TABLE II, above; SEQ ID NOs:61-63 correspond to the *ACTB* "control" gene

35   region used in the present analysis (*see* EXAMPLE 1, below)). Associated CpG island sequences are (based on the fact that the methylation state of a portion of a given CpG island

is generally representative of the island as a whole) those contiguous sequences of genomic DNA that encompass at least one nucleotide of the sequences defined by these specific oligonucleotide primers and probes, and satisfy the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.

5   These CpG island tags are then arrayed on solid supports (*e.g.*, nylon membranes, silicon, etc.), and probed with amplicons representing a pool of methylated CpG DNA, from test (*e.g.*, tumor) or reference samples. The differences in test and reference signal intensities on screened CpG island arrays reflect methylation alterations of corresponding sequences in the *test* DNA.

10   Comparison of the resulting data with the epigenetic patterns disclosed herein allows for a diagnostic or prognostic determination.

Therefore, according to this embodiment, pattern analysis (*see* working Examples 1-4, below) in a subset of CpG island tags, affixed to a solid support to form an array or microarray, is used to follow progression during various stages of cancer progression (*e.g.*,

15   gastrointestinal and esophageal dysplasia, gastrointestinal and esophageal metaplasia, Barrett's esophagous, and pre-cancerous conditions in normal esophageal squamus mucosa), and can be used to determine histological grades or stages of tumors, such as esophageal adenocarcinoma.

Other array or microarray embodiments of the present invention will be obvious to

20   those of ordinary skill in the relevant art. Such embodiments include, but are not limited to those wherein the specific primers and/or probes for *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2* and *TYMS* (*see* TABLES I and II, above; *and see* under "Definitions," above), corresponding to SEQ ID NOs:1-54, 58-60, 64 and 65 (*see* TABLE II, above; SEQ ID

25   NOs:61-63 correspond to the *ACTB* "control" gene region used in the present analysis (*see* EXAMPLE 1, above)) are arrayed on solid supports.

DISCUSSION

30   There is a need in the art for novel and more sensitive methods of cancer detection, chemoprediction and prognostics. There is a need in the art to define novel coordinate patterns of CpG island methylation changes (*i.e.*, novel epigenetic patterns) at multiple loci during progression of a disease, such as cancer. There is a need in the art to determine tumor-type-specific, and patient-specific epigenetic patterns or fingerprints. There is a need in the

35   art to provide biomarkers or probes, such as EAC-specific biomarkers or probes, that can be used in diagnostic and/or prognostic methods for the treatment of cancer. There is a need in

44

the art to determine whether esophageal adenocarcinoma displays a CIMP. There is a need in the art for novel methods for determining the stage of a tumor. The present invention addresses these needs.

A high-throughput, fluorescence-based methylation assay (MethyLight[TM]) was used
5   herein to examine and define novel hypermethylation patterns of 19 CpG islands and one non-CpG island during the progression of esophageal adenocarcinoma ("EAC"). The genes were thereby segregated into six classes of epigenetic patterns in the various tissue types. This is the most comprehensive methylation survey yet performed on a system having so many distinct histological stages of disease progression. Furthermore, the present analysis of
10   abnormal DNA hypermethylation offers a significant advantage over other approaches, such as gene expression analysis, in that it has greater sensitivity in the presence of contaminating normal cells, a common limiting factor.

DNA hypermethylation, as disclosed herein, is an early epigenetic alteration in the multi-step progression of EAC. The premalignant intestinal metaplasia ("IM," or Barret's
15   esophagus) is already significantly more methylated than the normal tissue (normal squamous mucosa). The present invention, in certain embodiments, provides the *novel* finding of frequent hypermethylation of five additional genes in this tumor system: *MYOD1, MGMT, CALCA, TIMP3,* and *HIC1.*

The methylation observed for *MGMT, TIMP3,* and *HIC1* in normal tissues may be
20   attributed to the particular region of the gene in which we analyzed methylation levels (Stoger et al., *Cell.* 73:61-71, 1993; Larsen et al., *Hum. Mol. Genet.* 2:775-80, 1993; Jones, P. A., *Trends Genet.* 15:34-37, 1999). These three genes were analyzed at CpG islands located at or downstream of the transcription start site (TABLE 2). However, this does not account for the *CALCA* methylation we observed, because we analyzed the promoter region of this gene.
25   Low levels of *CALCA* methylation has been previously reported in normal bone marrow samples of AML patients (Melki et al., *Cancer Res.* 59:3730-3740, 1999), suggesting that this locus may have a higher propensity to be methylated in normal tissues of cancer patients.

It is of particular interest to note that dysplastic tissues are more frequently methylated than stage I tumors for both Class A ($p < 0.0001$) and B ($p = 0.0174$) (Figure 1). This is
30   similar to the finding of genetic abnormalities (LOH, deletions and mutations) present in Barrett's esophagus with high grade dysplasia but not present in the adjacent invasive EAC (Barrett et al., *Nat. Genet.* 22:106-109, 1999). Because stage II-IV tumors appear to be methylated at Class A genes at a similar frequency as dysplasia, this suggests that stage I tumors may actually evolve from a different origin than the dysplastic tissue and higher
35   staged tumors, or may diverge after dysplasia independently from stage II-IV tumors during clonal expansion. Alternatively, but less likely, stage I tumors could undergo a transient

45

reversal of hypermethylation. Tumor development in Barrett's esophagus is proposed to evolve clonally through the linear multistep pathway of metaplasia-dysplasia-tumor (Zhuang et al., *Cancer Res.* 56:1961-4, 1996). However, the occurrence of genetic and, according to the present invention, epigenetic alterations in a non-linear order, indicates that the clonal

5 evolution of EAC is more complex than originally predicted (Barrett et al., *Nat. Genet.* 22:106-109, 1999). A similar observation has been described for different stages of bladder tumors (Salem et al., *Cancer Res.* 60:2473-2476, 2000).

There was, under the present analysis, no clear evidence, aside from one tumor with 10 genes methylated, for a separate cluster of tumors with extensive concordant methylation,

10 indicative of a CpG island methylator phenotype ("CIMP"). Similar results were obtained even when only "type-C" genes, as defined for CIMP (methylated in cancer, not methylated in normal tissues; Toyota et al., *Proc. Natl. Acad. Sci. USA* 96:8681-8686, 1999; Toyota et al., *Cancer Res.* 59:5438-5442, 1999), were examined. Interestingly, the "type-C" genes in EAC differ from those described for colorectal cancer (*Id*). For example, *ESR1* is classified as a

15 "type-A" (defined as methylated in aging normal tissues) rather than a "type-C" gene in colorectal cancer, because it is frequently methylated in the normal colonic epithelium of aging individuals (*Id*). However, in esophageal adenocarcinoma, *ESR1* clearly behaves as a "type-C" gene. This may be attributed to the difference in the technology used to measure hypermethylation, or more likely may be due to differences in tissue types.

20 According to the present invention, there is a tissue-specific and tumor-specific propensity for particular genes to become hypermethylated. For instance, *APC* is hypermethylated in normal stomach, but not in normal esophageal mucosa. The tumor-specificity of hypermethylation is illustrated by the lack of detectable methylation of the two Class E genes *TGFBR2* and *RB1*, which are frequently hypermethylated in gastric and lung

25 tumors, and retinoblastoma tumors, respectively (Stirzaker et al., *Cancer Res.* 57:2229-2237, 1997; Kang et al., *Oncogene* 18:7280-7286, 1999; Hougaard et al., *Br. J. Cancer* 79:1005-1011, 1999).

The tumor-specificity of CpG island hypermethylation suggests that there may be tissue-specific trans-acting factors that modulate methylation changes of these CpG islands

30 during tumorigenesis and which differ between esophageal adenocarcinomas and other tumor types. Alternatively, there may be a lack of selective advantage to the silencing of these genes in esophageal adenocarcinomas by DNA methylation. There are two scenarios in which this would be the case. One is if the gene in question has been inactivated by a different, genetic mechanism, rendering hypermethylation of no further selective advantage. The other is if the

35 gene does not play a role in tumor suppression in this particular tumor system.

Although alterations in DNA methylation changes are common events in

46

tumorigenesis, the underlying mechanism is unclear. Abnormal methylation, at least in colorectal tumors, is not due to a mere upregulation of the DNA methyltranseferase genes, suggesting that other major players are involved (Eads et al., *Cancer Res*. 59:2302-2306, 1999). The present invention provides some first glimpses into the process underlying these
5   abnormal methylation changes.

According to the present invention, different, functionally unrelated, genes can behave in distinct classes with respect to their methylation changes within various tissues of EAC progression. The CpG island hypermethylation does not appear to be a random, stochastic process (although there is a stochastic component), but rather a step-wise process that
10   involves multiple, distinct groups of alterations. This is consistent with the existence of several different mechanisms that protect against CpG island hypermethylation. In this scenario, the concerted changes seen at different CpG islands would be the result of the loss of a different type of protective element at different stages of disease progression. This finding does not appear to be dependent on the location of the CpG island relative to the gene,
15   since both promoter and internal CpG islands were observed in all gene classes. The structural features of these CpG islands were also examined under the present analysis by analyzing the %GC content, the observed/expected CpG ratio and the CpG:GpC ratio and found no association with gene class (TABLE 2).

According to the present invention, the IM or NE samples themselves, with or without
20   associated dysplasia or cancer, were histologically indistinguishable, yet molecularly distinct. NE and IM samples derived from individuals with concurrent distally located dysplasia or malignancy show a statistically higher incidence of CpG island hypermethylation. These findings were confirmed herein in the IM tissues in a completely independent study. This provides strong support for the use of epigenetic markers, particularly Class A and B genes, as
25   disease screening tools and as predictive markers for the progression of more advanced staged disease.

The methylation profiles of the present invention provide methods and compositions for the early detection of cancer. Such a molecular diagnostic approach using normal and/or premalignant tissues to identify patients with cancer or at elevated risk for developing cancer
30   provides an opportunity for early intervention. Furthermore, a benefit of using CpG island hypermethylation as a diagnostic or prognostic marker is that it can easily be detected in a field of normal cell contamination as a gain of signal, unlike loss of gene expression (*e.g.*, LOH and deletion analysis), which is difficult to resolve in a sample with contaminating normal cells.

47

SUMMARY

According to the present invention, the 19 CpG islands (TABLES I and II) studied segregate into six classes of epigenetic patterns in the various tissue types. Each class
5   undergoes unique epigenetic changes at different steps of disease progression of EAC. The methylation profiles provide methods and compositions for the early detection of cancer.

I claim:

     1.   A method for diagnosing cancer or cancer-related conditions from tissue samples, comprising:

5   (a)   obtaining a tissue sample from a test tissue or region to be diagnosed;

     (b)   performing a methylation assay of the tissue sample, wherein the methylation assay determines the methylation state of genomic CpG sequences, wherein the genomic CpG sequences are located within at least one gene sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1,*

10 *MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR,* and combinations thereof; and

     (c)   making a diagnostic or prognostic prediction of the cancer based, at least in part, upon the methylation state of the genomic CpG sequences.

     2.   The method of claim 1, wherein the genomic CpG sequences located within at

15 least one gene sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2* and *TYMS,* correspond to genomic CpG sequences of CpG islands.

     3.   The method of claim 1, wherein the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3,*

20 *CTNNB1, PTGS2, TYMS* and *MTHFR* gene sequences are those defined by the specific oligonucleotide primers and probes corresponding to SEQ ID Nos:1-60, 64 and 65, as listed in TABLE II, or portions thereof.

     4.   The method of claim 2 wherein the CpG islands are located within the promoter regions of one or more of the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B,*

25 *ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2* and *TYMS* genes.

     5.   The method of claim 2, wherein the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2,* and *TYMS* gene sequences correspond to any CpG island sequences

30 associated with the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:1-54, 58-60, 64 and 65, as listed in TABLE II, or portions thereof, and wherein the associated CpG island sequences are those contiguous sequences of genomic DNA that encompass at least one nucleotide of the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:1-54, 58-60, 64 and 65,

35 and satisfy the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.

6.   The method of claim 1, wherein the genomic CpG sequences are located within at least one gene sequence selected from the group consisting of *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3*, and combinations thereof.

7.   The method of claim 6, wherein the genomic CpG sequences located within at least one gene sequence selected from the group consisting of *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3*, correspond to genomic CpG sequences of CpG islands.

8.   The method of claim 6, wherein the *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3* gene sequences are those defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:19-21, SEQ ID NOs:1-3, SEQ ID NOs:7-9, SEQ ID NOs:10-12, SEQ ID NOs:4-6, SEQ ID NOs:16-18 and SEQ ID NOs:13-15, respectively, as listed in TABLE II.

9.   The method of claim 7 wherein the CpG islands are located within the promoter regions of one or more of the *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3* genes.

10.   The method of claim 7 wherein the *APC, CDKN2A, MYODI, CALCA, ESRI, MGMT* and *TIMP3* gene sequences correspond to any CpG island sequences associated with the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:19-21, SEQ ID NOs:1-3, SEQ ID NOs:7-9, SEQ ID NOs:10-12, SEQ ID NOs:4-6, SEQ ID NOs:16-18 and SEQ ID NOs:13-15, respectively, as listed in TABLE II, or portions thereof, and wherein the associated CpG island sequences are those contiguous sequences of genomic DNA that encompass at least one nucleotide of the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:19-21, SEQ ID NOs:1-3, SEQ ID NOs:7-9, SEQ ID NOs:10-12, SEQ ID NOs:4-6, SEQ ID NOs:16-18 and SEQ ID NOs:13-15, and satisfy the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.

11.   The method of claim 1, wherein the cancer or cancer-related condition is selected from the group consisting of gastrointestinal or esophageal adenocarcinoma, gastrointestinal or esophageal dysplasia, gastrointestinal or esophageal metaplasia, Barrett's intestinal tissue, pre-cancerous conditions in normal esophageal squamous mucosa, and combinations thereof.

12.   The method of claim 11, wherein the cancer is esophageal adenocarcinoma, and wherein making a diagnostic or prognostic prediction of the cancer, based upon the methylation state of the genomic CpG sequences provides for classification of the adenocarcinoma by grade or stage.

13.   The method of claim 6, wherein the cancer or cancer-related condition is selected from the group consisting of gastrointestinal or esophageal adenocarcinoma,

50

gastrointestinal or esophageal dysplasia, gastrointestinal or esophageal metaplasia, Barrett's intestinal tissue, pre-cancerous conditions in normal esophageal squamous mucosa, and combinations thereof.

14.   The method of claim 13, wherein the cancer is esophageal adenocarcinoma, and wherein making a diagnostic or prognostic prediction of the cancer, based upon the methylation state of the genomic CpG sequences provides for classification of the adenocarcinoma by grade or stage.

15.   The method of claim 1, wherein the methylation assay used to determine the methylation state of genomic CpG sequences is selected from the group consisting of MethylLight™, MS-SNuPE, MSP, COBRA, MCA, and DMH, and combinations thereof.

16.   The method of claim 6, wherein the methylation assay used to determine the methylation state of genomic CpG sequences is selected from the group consisting of MethylLight™, MS-SNuPE, MSP, COBRA, MCA and DMH, and combinations thereof.

17.   The method of claim 1, wherein the methylation assay used to determine the methylation state of genomic CpG sequences is based, at least in part, on an array or microarray comprising CpG-containing sequences located within at least one gene sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR*.

18.   The method of claim 17, wherein the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2,* and *TYMS* gene sequences correspond to any CpG island sequences associated with the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:1-54, 58-60, 64 and 65, as listed in TABLE II, or portions thereof, and wherein the associated CpG island sequences are those contiguous sequences of genomic DNA that encompass at least one nucleotide of the sequences defined by the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:1-54, 58-60, 64 and 65, and satisfy the criteria of having both a frequency of CpG dinucleotides corresponding to an Observed/Expected Ratio >0.6, and a GC Content >0.5.

19.   The method of claim 17, wherein the *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR* gene sequences are those defined by, or correspond to the specific oligonucleotide primers and probes corresponding to SEQ ID NOs:1-60, 64 and 65, as listed in TABLE II, or portions thereof.

20.   The method of claim 1 wherein the methylation state of genomic CpG sequences that is determined is that of hypermethylation, hypomethylation or normal

methylation.

21.   A kit useful for diagnosis or prognosis of cancer or cancer-related conditions, comprising a carrier means containing one or more containers comprising:

(a)   a container containing a probe or primer which hybridizes to any region of a sequence located within at least one gene sequence selected from the group consisting of *APC, ARF, CALCA, CDH1, CDKN2A, CDKN2B, ESR1, GSTP1, HIC1, MGMT, MLH1, MYOD1, RB1, TGFBR2, THBS1, TIMP3, CTNNB1, PTGS2, TYMS* and *MTHFR*; and

(b)   additional standard methylation assay reagents required to affect detection of methylated CpG-containing nucleic acid based, at least in part, on the probe or primer.

22.   The kit of claim 21, wherein the additional standard methylation assay reagents are standard reagents for performing a methylation assay from the group consisting of MethyLight™, MS-SNuPE, MSP, COBRA, MCA and DMH, and combinations thereof.

23.   The kit of claim 21, wherein the probe or primer comprises at least about 12 to 15 nucleotides of a sequence selected from the group consisting of SEQ ID NOs:1-60, 64 and 65, as listed in TABLE II.

24.   A kit useful for diagnosis or prognosis of cancer or cancer-related conditions, comprising a carrier means containing one or more containers comprising:

(a)   an array or micorarray comprising sequences of at least about 12 to 15 nucleotides of a sequence selected from the group consisting of SEQ ID NOs:1-60, 64, 65, and any sequence located within a CpG island sequence associated with SEQ ID NOs:1-54, 58-60, 64 and 65.

Fig. 1

Fig. 2

## MEAN PERCENT OF GENES METHYLATED
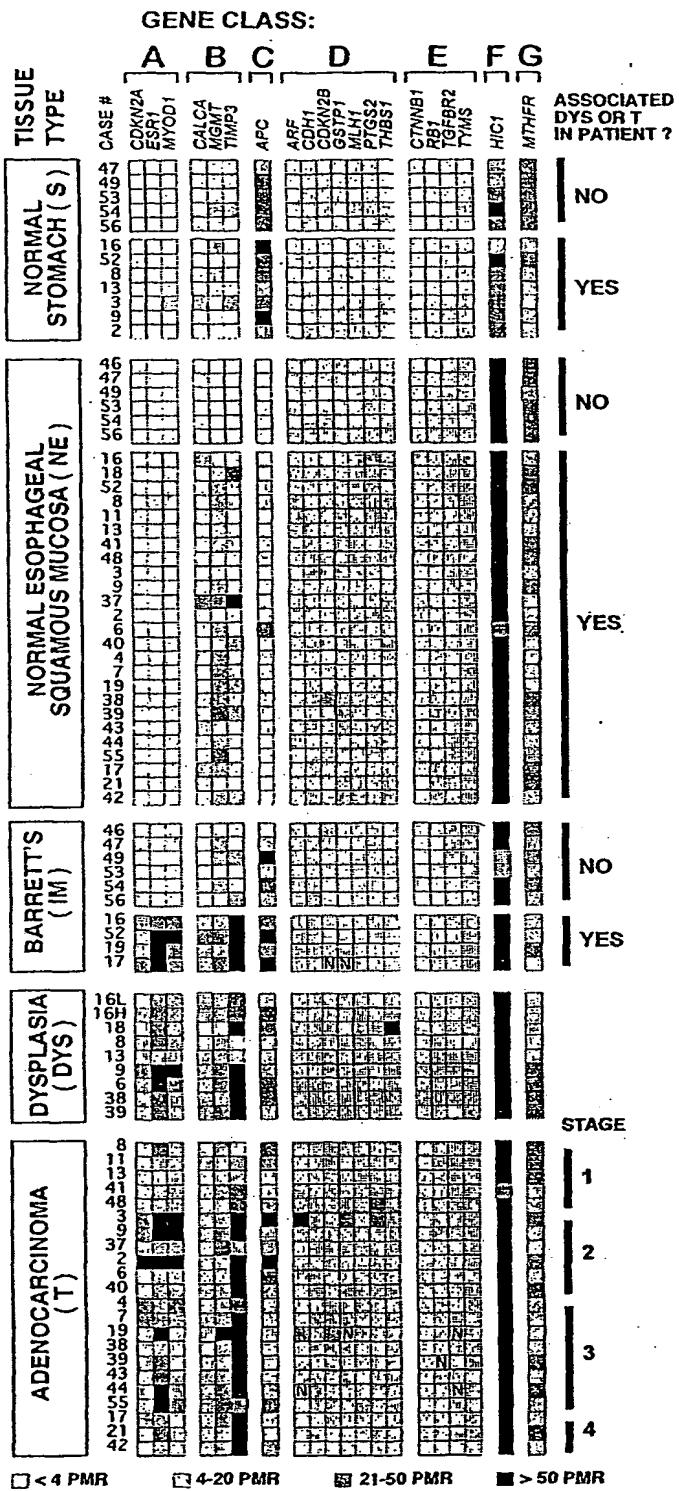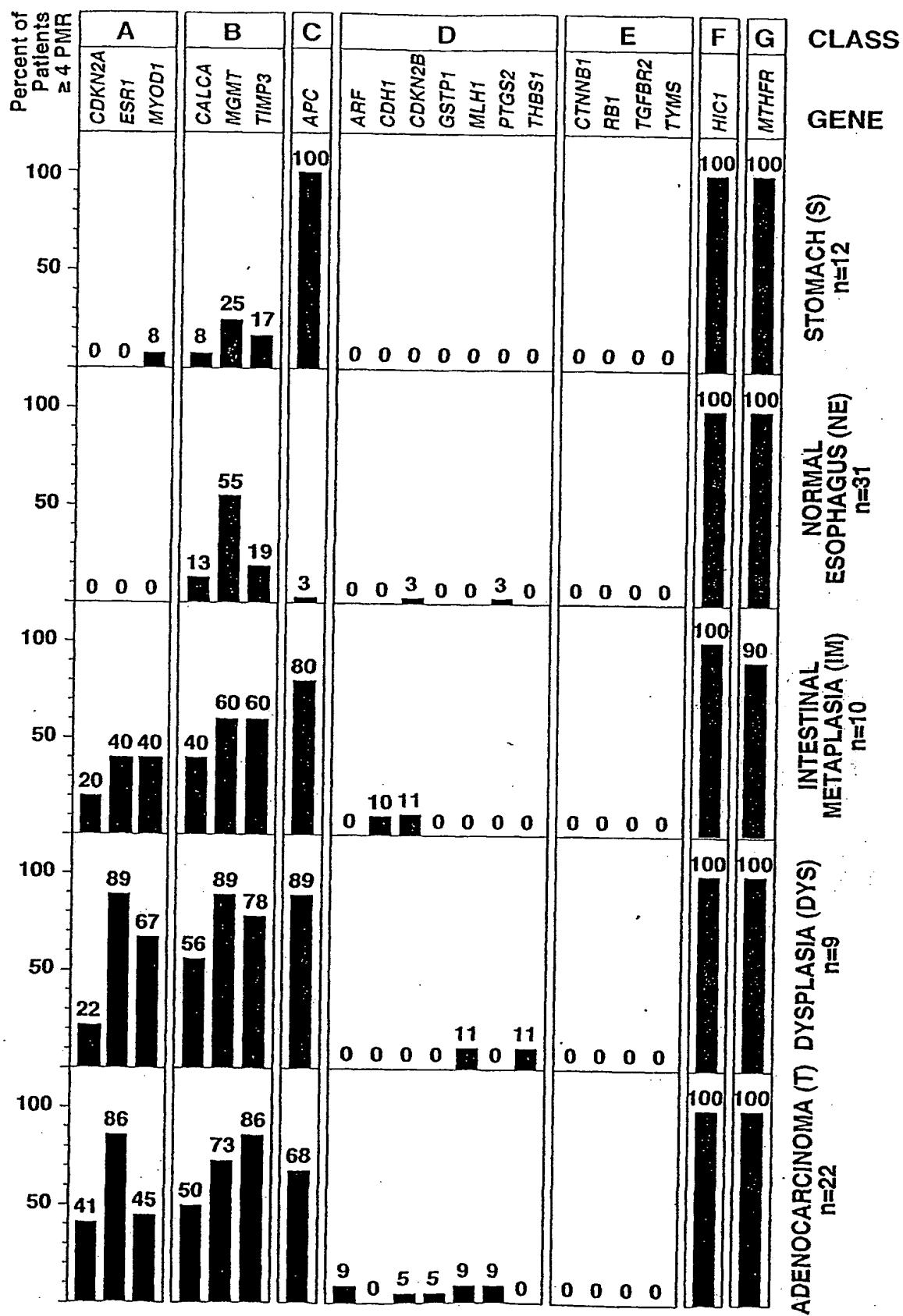
GENE CLASS — S, NE, IM, DYS, T
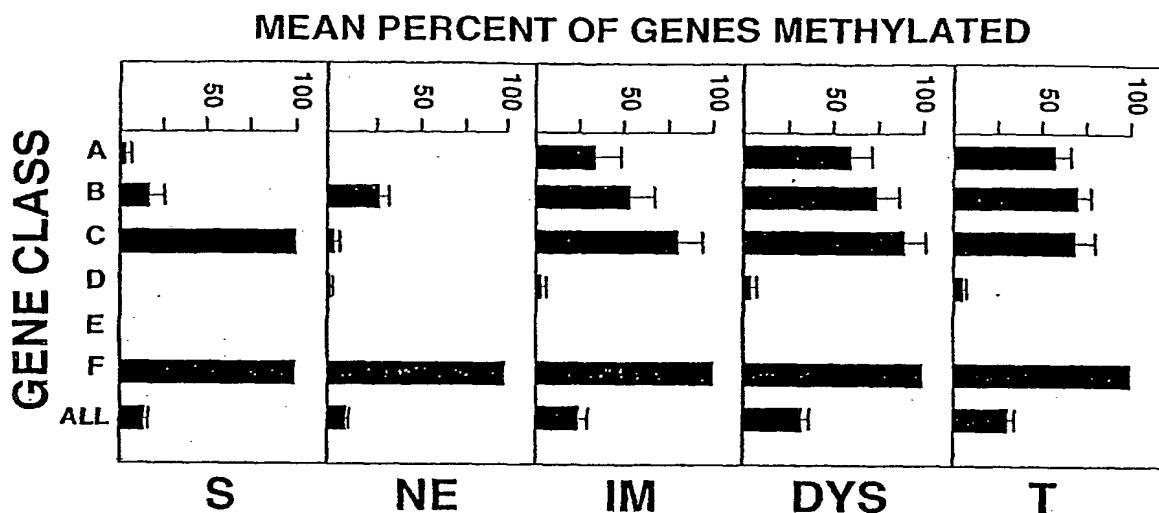
TISSUE COMPARISON

p-values	NE → IM	IM → DYS	IM → T	DYS → T
A	0.0006	0.031	0.015	0.87
B	0.027	0.13	0.15	0.71
C	< 0.0001	0.55	0.34	0.11
D	0.46	0.92	0.40	0.48
E	—	—	—	—
F	—	—	—	—
ALL	< 0.0001	0.044	0.042	0.70
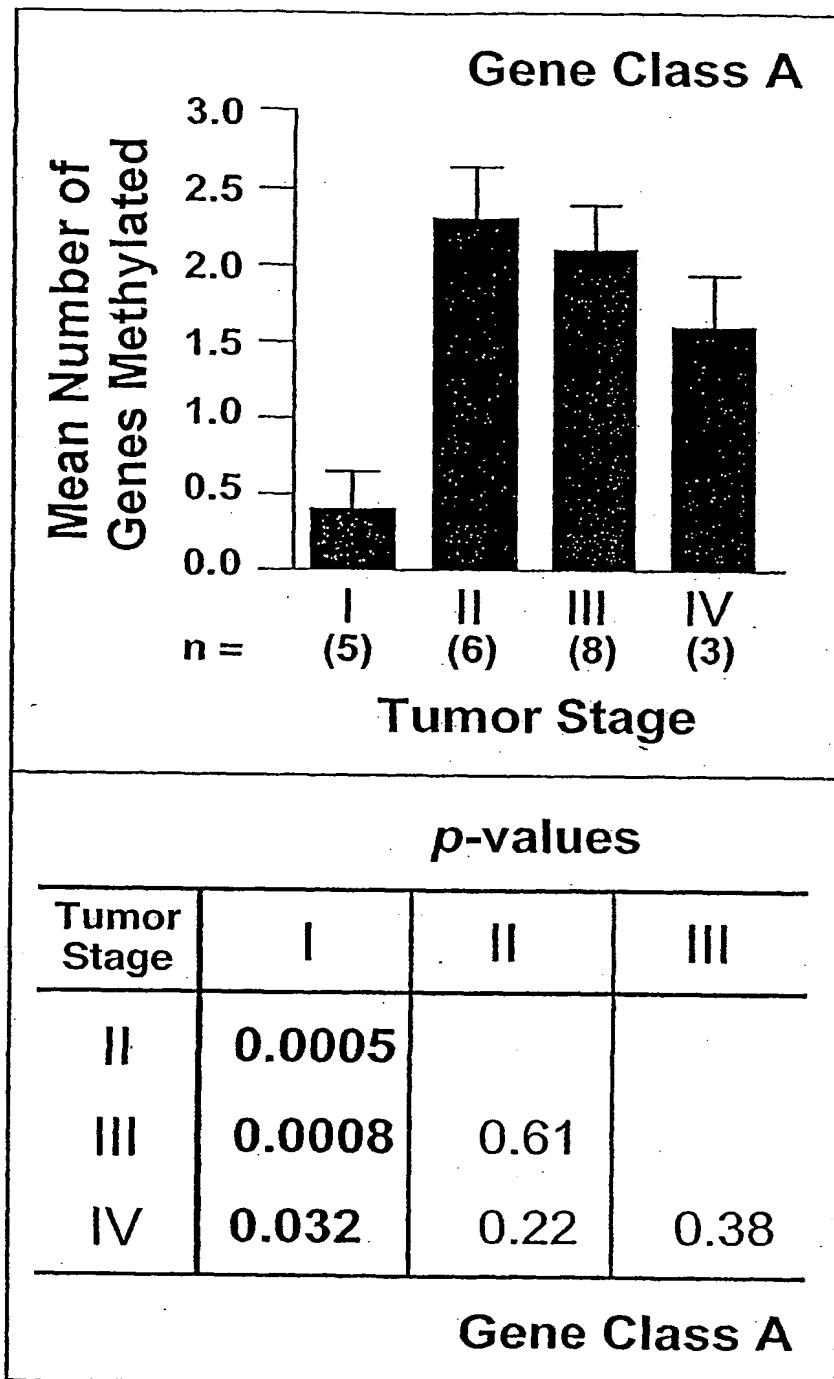
GENE CLASS

Fig. 3

Fig. 4
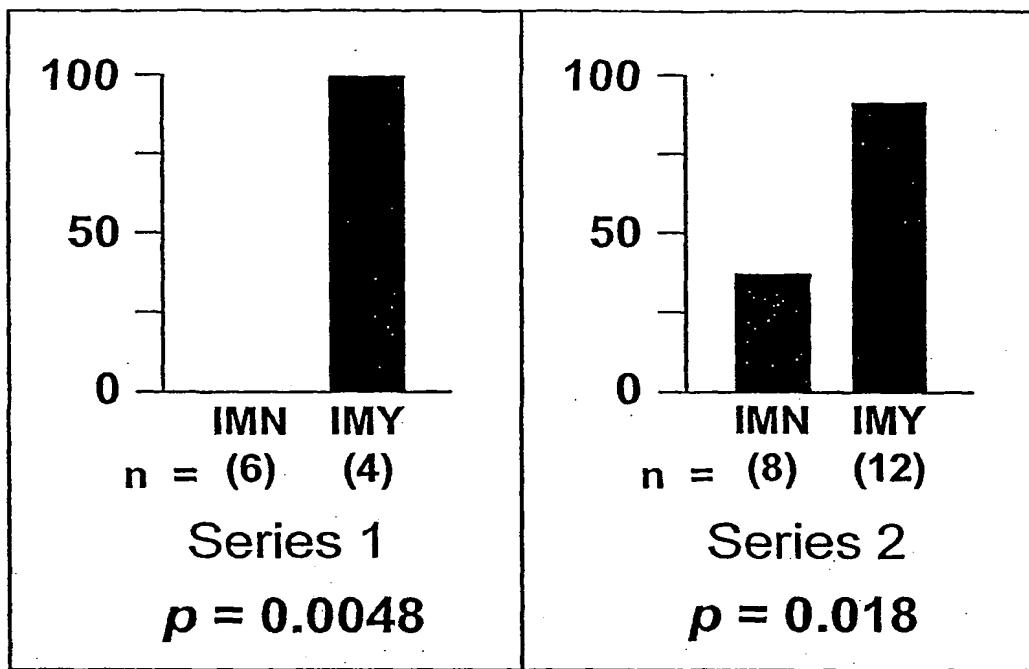
INTESTINAL METAPLASIA (IM)
(IMY = Associated Dysplasia)
(IMN = No Associated Dysplasia)

Fig. 5

Fig. 6

BEST AVAILABLE COPY

SEQUENCE LISTING

<110> LAIRD, Peter
      EADS, Cindy

<120>   EPIGENETIC SEQUENCES FOR ESOPHAGEAL ADENOCARCINOMA

<130>   47675-12

<140>   60/193,839
<141>   2000-03-31

<160>   65

<170>   PatentIn version 3.0

<210>   1
<211>   22
<212>   DNA
<213>   Homo sapiens

<400>   1
tggaattttc ggttgattgg tt

22

<210>   2
<211>   19
<212>   DNA
<213>   Homo sapiens

<400>   2
aacaacgtcc gcacctcct

19

<210>   3
<211>   18
<212>   DNA
<213>   Homo sapiens

```
<400>    3
acccgacccc gaaccgcg
```

18


```
<210>    4
<211>    19
<212>    DNA
<213>    Homo sapiens
```

```
<400>    4
ggcgttcgtt ttgggattg
```

19


```
<210>    5
<211>    19
<212>    DNA
<213>    Homo sapiens
```

```
<400>    5
gccgacacgc gaactctaa
```

19


```
<210>    6
<211>    24
<212>    DNA
<213>    Homo sapiens
```

```
<400>    6
cgataaaacc gaacgacccg acga
```

24


```
<210>    7
<211>    18
<212>    DNA
<213>    Homo sapiens
```

<400>   7
gagcgcgcgt agttagcg

18 .

<210>   8
<211>   17
<212>   DNA
<213>   Homo sapiens

<400>   8
tccgacacgc cctttcc

17

<210>   9
<211>   30
<212>   DNA
<213>   Homo sapiens .

<400>   9
ctccaacacc cgactactat atccgcgaaa

30

<210>   10
<211>   23
<212>   DNA
<213>   Homo sapiens

<400>   10
gttttggaag tatgagggtg acg

23

<210>   11
<211>   19
<212>   DNA

<213>   Homo sapiens

<400>   11
ttcccgccgc tataaatcg

19

<210>   12
<211>   30
<212>   DNA
<213>   Homo sapiens

<400>   12
attccgccaa tacacaacaa ccaataaacg

30

<210>   13
<211>   21
<212>   DNA
<213>   Homo sapiens

<400>   13
gcgtcggagg ttaaggttgt t

21

<210>   14
<211>   22
<212>   DNA
<213>   Homo sapiens

<400>   14
ctctccaaaa ttaccgtacg cg

22

<210>   15
<211>   19

4

<212>   DNA
<213>   Homo sapiens

<400>   15
aactcgctcg cccgccgaa

19

<210>   16
<211>   28
<212>   DNA
<213>   Homo sapiens

<400>   16
ctaacgtata acgaaaatcg taacaacc

28

<210>   17
<211>   25
<212>   DNA
<213>   Homo sapiens

<400>   17
agtatgaagg gtaggaagaa ttcgg

25

<210>   18
<211>   30
<212>   DNA
<213>   Homo sapiens

<400>   18
ccttacctct aaataccaac cccaaacccg

30

<210>   19

<211> 19
<212> DNA
<213> Homo sapiens

<400> 19
gaaccaaaac gctccccat

19

<210> 20
<211> 27
<212> DNA
<213> Homo sapiens

<400> 20
ttatatgtcg gttacgtgcg tttatat

27

<210> 21
<211> 22
<212> DNA
<213> Homo sapiens

<400> 21
cccgtcgaaa acccgccgat ta

22

<210> 22
<211> 19
<212> DNA
<213> Homo sapiens

<400> 22
acgggcgttt tcggtagtt

19

6

```
<210>   23
<211>   20
<212>   DNA
<213>   Homo sapiens

<400>   23
ccgaacctcc aaaatctcga


20



<210>   24
<211>   26
<212>   DNA
<213>   Homo sapiens

<400>   24
cgactctaaa ccctacgcac gcgaaa


26



<210>   25
<211>   26
<212>   DNA
<213>   Homo sapiens

<400>   25
aattttaggt tagagggtta tcgcgt


26



<210>   26
<211>   22
<212>   DNA
<213>   Homo sapiens

<400>   26
tccccaaaac gaaactaacg ac


22
```

```
<210>   27
<211>   19
<212>   DNA
<213>   Homo sapiens

<400>   27
cgcccacccg acctcgcat

19



<210>   28
<211>   20
<212>   DNA
<213>   Homo sapiens

<400>   28
aggaaggaga gagtgcgtcg

20



<210>   29
<211>   21
<212>   DNA
<213>   Homo sapiens

<400>   29
cgaataatcc accgttaacc g

21



<210>   30
<211>   29
<212>   DNA
<213>   Homo sapiens

<400>   30
ttaacgacac tcttcccttc tttcccacg

29
```

```
<210>   31
<211>   23
<212>   DNA
<213>   Homo sapiens

<400>   31
gtcggcgtcg tgatttagta ttg

23


<210>   32
<211>   23
<212>   DNA
<213>   Homo sapiens

<400>   32
aaactacgac gacgaaactc caa

23


<210>   33
<211>   29
<212>   DNA
<213>   Homo sapiens

<400>   33
aaacctcgcg acctccgaac cttataaaa

29


<210>   34
<211>   18
<212>   DNA
<213>   Homo sapiens

<400>   34
ctatcgccgc ctcatcgt
```

18

<210>   35
<211>   30
<212>   DNA
<213>   Homo sapiens

<400>   35
cgttatatat cgttcgtagt attcgtgttt

30

<210>   36
<211>   22
<212>   DNA
<213>   Homo sapiens

<400>   36
cgcgacgtca aacgccacta cg

22

<210>   37
<211>   19
<212>   DNA
<213>   Homo sapiens

<400>   37
cggaagcgtt cgggtaaag

19

<210>   38
<211>   18
<212>   DNA
<213>   Homo sapiens

<400>   38
aattccaccg ccccaaac

18

<210>   39
<211>   29
<212>   DNA
<213>   Homo sapiens

<400>   39
tttccgccaa atatcttttc ttcttcgca

29

<210>   40
<211>   18
<212>   DNA
<213>   Homo sapiens

<400>   40
cgacgcacca acctaccg

18

<210>   41
<211>   25
<212>   DNA
<213>   Homo sapiens

<400>   41
gttttgagtt ggttttacgt tcgtt

25

<210>   42
<211>   19
<212>   DNA
<213>   Homo sapiens

<400>   42

acgccgcgct cacctccct

19

<210>   43
<211>   17
<212>   DNA
<213>   Homo sapiens

<400>   43
ggaaaggcgc gtcgagt

17

<210>   44
<211>   18
<212>   DNA
<213>   Homo sapiens

<400>   44
tcccctatcc caaacccg

18

<210>   45
<211>   18
<212>   DNA
<213>   Homo sapiens

<400>   45
cgcgcgtttc ccgaaccg

18

<210>   46
<211>   22
<212>   DNA
<213>   Homo sapiens

```
<400>  46
ttagttcgcg tatcgattag cg

22


<210>  47
<211>  18
<212>  DNA
<213>  Homo sapiens

<400>  47
actaaacgcc gcgtccaa

18


<210>  48
<211>  21
<212>  DNA
<213>  Homo sapiens

<400>  48
tcacgtccgc gaaactcccg a

21


<210>  49
<211>  18
<212>  DNA
<213>  Homo sapiens

<400>  49
gcgcggagcg tagttagg

18


<210>  50
<211>  20
<212>  DNA
<213>  Homo sapiens
```

```
<400>   50
caaaccccgc tactcgtcat
```

20

```
<210>   51
<211>   21
<212>   DNA
<213>   Homo sapiens
```

```
<400>   51
cacgaacgac gccttcccga a
```

21

```
<210>   52
<211>   19
<212>   DNA
<213>   Homo sapiens
```

```
<400>   52
cggcgttagg aaggacgat
```

19

```
<210>   53
<211>   24
<212>   DNA
<213>   Homo sapiens
```

```
<400>   53
tctcaaacta taacgcgcct acat
```

24

```
<210>   54
<211>   29
<212>   DNA
```

```
<213>   Homo sapiens

<400>   54
ccgaataccg acaaaatacc gatacccgt

29


<210>   55
<211>   29
<212>   DNA
<213>   Homo sapiens

<400>   55
tggtagtgag agttttaaag atagttcga

29


<210>   56
<211>   18
<212>   DNA
<213>   Homo sapiens

<400>   56
cgcctcatct tctcccga

18


<210>   57
<211>   27
<212>   DNA
<213>   Homo sapiens

<400>   57
tctcataccg ctcaaaatcc aaacccg

27


<210>   58
<211>   19
```

```
<212>  DNA
<213>  Homo sapiens

<400>  58
gttaggcggt tagggcgtc

19


<210>  59
<211>  19
<212>  DNA
<213>  Homo sapiens

<400>  59
ccgaacgcct ccatcgtat

19


<210>  60
<211>  31
<212>  DNA
<213>  Homo sapiens

<400>  60
caacatcgtc tacccaacac actctcctac g

31


<210>  61
<211>  25
<212>  DNA
<213>  Homo sapiens

<400>  61
tggtgatgga ggaggtttag taagt

25


<210>  62
```

<211>   27
<212>   DNA
<213>   Homo sapiens

<400>   62
aaccaataaa acctactcct cccttaa

27

<210>   63
<211>   30
<212>   DNA
<213>   Homo sapiens

<400>   63
accaccaccc aacacacaat aacaaacaca

30

<210>   64
<211>   22
<212>   DNA
<213>   Homo sapiens

<400>   64
tggagttttc ggttgattgg tt

22

<210>   65
<211>   19
<212>   DNA
<213>   Homo sapiens

<400>   65
aacaacgccc gcacctcct

19

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12Q 1/68
US CL :435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6: 536/24.31, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	KAWAKAMI, K et al. Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. Journal of the National Cancer Institute. 15 November 2000. vol. 92., No. 22, pages 1805-1811, see entire document.	1-24
Y,P	EADS, C.A. et al. Fields of Aberrant CpG Island Hypermethylation in Barrett's Esophagus and associated Adenocarcinoma. Cancer Research. 15 September 2000. Vol. 50, pages 5021-5026, see entire document.	1-24
Y	SUZUKI, H. et al. Intragenic mutations of CDKN2B and CDKN2A iHn Primary human esophageal cancers. Human Molecular Genetics. 1995. Vol. 4. no. 10, pages 1883-1887, see entire document.	1-24

☒ Further documents are listed in the continuation of Box C.	☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
"O"	document referring to an oral disclosure, use, exhibition or other means	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 JULY 2001	24 AUG 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer LISA ARTHUR
Facsimile No. (703) 305-3230	Telephone No. (703) 308-3988
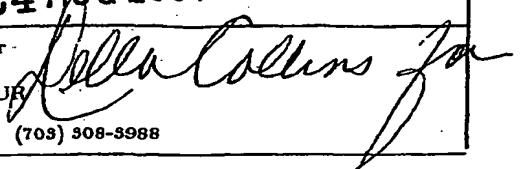
Form PCT/ISA/210 (second sheet) (July 1998)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CODY, D.T, ii et al. Differential DNA methylation of the p16 INK4A/CDKN2A promoter in human opral cacner cells and normal human oral karatinocytes. ORAL ONCOLOGY. 1999, Vol. 35, pages 516-522, see entire document.	1-10, 13, 15-24
Y	GRAFF, J.R. et al. Distinct patterns of E-Cadherin CpG island methylation in papillary, follicular, Hurtle's cell and poorly differentiated thyroid carcinoma. Cacner Research. 15 May 1998, Vol. 58, pages 2063-2066, see entire document.	1-10,15-24
Y	IWASE, H. et al. DNA methylation analysis at distal and proximal promoter regions of the oestrogen receptor gene in breast cancers. British Journal of Cancer. 1999, Vol. 80. No. 12, pages 1982-1986, see entire document.	1-10, 15-24
Y	MILLAR, D. S. et al. Detailed methylation analysis of the glutathione S-transferase(GSTP1) gene in prostate cancer. ONCOGENE. 1999,Vol. 18. pages 1313-1324, see entire document.	1-10, 15-24
Y	JHAVERI, M.S. et al. Methylation-mediated regulation of the glutathione S-transferase Pi gene in human breast cance r cells. Gene (1998) Vol. 210, pages 1-7, see entire document.	1-10 and 15-24
Y	BARRETT, M.T. et al. Evolution of neoplastic cell lineages in Barrett oseophagus. Nature Genetics. May 1999, Vol. 22. pages 106-109, see entire document.	1-24
Y	HERMAN, J.G. et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proceedings of the National Academy of Sciences. June 1998. vol. 95. pages 6870-6875, see entire document.	1-11, 13, 15-24
Y	ESTELLER, M et al. Inactivation of the DNA Repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. Cancer Research, 15 February 1999. Vol. 59. pages 793-797, see entire document.	1-24
Y	IACOPETTA, B.J. et al. Hypermethylation of the Myf-3 gene in human colorectal cancer. Anticancer Research. 1997. Vol. 17. pages 429-432, see entire document.	1-11, 13, 15-24
Y	ESTELLER, M. et al. Hypermethylation-associated inactivation of p14ARF is indepdendent of p16INK4A methylation and p53 mutational status. Cancer Research. 01 January 2000. Vol. 60, pages 128-133, see entire document.	1-24

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	NAKAMURA, M. et al. Promoter hypermethylation of the RB1 gene in glioblastomas. Laboratory Investigations. January 2001. Vol. 81. no. 1, page 77-82, see entire document.	1-10, 15-24
Y	MEKI, J.R. et al. Cancer-specific region of hypermethylation identified within the HIC1 putative tumour suppressor gene in acute myeloid leukaemia. Leukemia. 1999. Vol. 13. pages 877-883, see entire document.	1-10, 15-24

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

WEST, MEDLINE, BIOSIS, EMBASE, CAPLUS
search terms: hypermethylation, cpg islands, apc, arf,calca, cdh1, cdkn2a, cdkn2b, esr1, gstp1, hic1, mgmt, mlh1, mydo1, rb1, tgtbr2, thbs1, timp3, ctnnb1m ptgs2, tyms, mthfr